Beyond AlphaGo


DATE & LOCATION
October 27-28, 2016
Grand Hall(5F), aT Center, Seoul, Korea
Gangnam-daero 27, Seocho-gu, Seoul

ORGANIZERS
Hosted by Korea Ministry of Science, ICT, and Future Planning

Sponsored by Korean Institute of Information Scientists and Engineers (KIISE)


Patrons Institute for Information and Communications Technology Promotion, ETRI, Samsung, LG, SK Telecom, Naver, Doosan, POSCO, Google Korea, Microsoft Korea, Qualcomm Korea, Nvidia Korea, Robotis, Yujin Robot, Future Robot, Surromind Robotics, Seoul National University Institute of Cognitive Science, Cognitive Robotics and Artificial Intelligence Center, Yonsei Cognitive Engineering Square, Mirae Asset Financial Group, Shinhan Financial Group, KB Financial Group, Deloitte
CONTENTS

AIMS AND SCOPE .............................................................................................................................................4

COMMITTEES .....................................................................................................................................................5

PROGRAM ...........................................................................................................................................................6

- Instructible Social Robots : Matthias Scheutz ...................................................................................................7
- Robot Intelligence vs Biological Intelligence : Aaron Sloman ..........................................................................8
- Natural and Artificial Intelligence: Towards Neuromorphic Computational Systems : Klaus Mainzer ..........9
- Digital Companions for Seamless Co-Creation : Andreas Dengel ..................................................................10
- A Bayesian Perspective of Intelligence : Hideki Asoh .....................................................................................11

POSTER INDEX .................................................................................................................................................12

REGISTRATION ..................................................................................................................................................16

VENUE ................................................................................................................................................................17

FLOOR PLAN ......................................................................................................................................................18

POSTER SESSION ............................................................................................................................................19
The International Symposium on Perception, Action, and Cognitive Systems (PACS) is a premier venue for science and engineering for embodied cognitive systems – natural and artificial - that sense, act, feel, think, reason, communicate, learn and evolve in real-world environments. The fundamental significance of embodied cognitive systems has long been recognized in science, but its industrial importance has realized only recently with new technologies such as the Internet of things, mixed reality, wearable devices, personal robots, and autonomous cars. The goal of PACS is to bring international researchers from academia and industry together to present recent progresses and discuss new frontiers in interdisciplinary research and technology for embodied cognitive systems.

The topics of PACS cover the following (not exhaustive):

- Action Science
- Anticipatory Systems
- Artificial Intelligence
- Augmented Cognition
- Autonomous Learning
- Brain-like Systems
- Cognitive Architectures
- Cognitive Robots
- Complex Adaptive Systems
- Conversational Agents
- Cyber-Physical Systems
- Embodied Cognition
- Emotion Machines
- Haptic Interfaces
- Human-Robot Interaction
- Human-Level AI
- Intelligence Augmentation
- Internet of Things
- Machine Cognition
- Mind Machines
- Perceptual Computing
- Personal Robots
- Self-aware Systems
- Smart Machines
- Wearable Agents
Advisory Committee
Chungmin Lee (SNU)
Soo-Young Lee (KAIST)
Jin-Hyung Kim (AII)
Tomaso Poggio (MIT)
II-Hong Suh (Hanyang Univ.)
Ipke Wachsmuth (CITEC)
Wolfgang Wahlster (DFKI)

General Co-Chairs
Choong Seon Hong (KIIE & KHU)
Klaus Mainzer (TU Munich)

Program Committee
Co-Chairs
Andreas Dengel (DFKI)
Byoung-Tak Zhang (SNU)

Members
Seong-Bae Cho (Yonsei)
Byung-Wook Chun (HMC)
Sung Il Chung (Deloitte)
Dayk Jang (SNU)
June Hyun Jang (Doosan)
Da-Shik Kim (KAIST)
Yong-Hwan Kim (LG U+)
Geunbae Lee (Samsung)
Kang-Young Lee (IBM)
Kyoung Lee (Saltlux)
Min-Ho Lee (KNU)
Miran Lee (MSR)
Sang-Hyun Lee (POSOC)
Te-Won Lee (Qualcomm)
Yong-Hoon Lee (Humax)
Chaeduk Lim (ETRI)
Myung-Soon Park (SK Telecom)
Sang-Kyu Park (ETRI)
Seongil Seo (MSIP)
Kyung-Cheol Shin (Yujin Robot)
Chang Song (Naver Labs)

Organizing Committee
Co-Chairs
Hideki Asob (AIRC)
Kwangsu Cho (Yonsei Univ.)

Members
Jae-Shik Choi (UNIST)
Jung-Hwan Choi (NIA)
Jung-Hyun Han (Korea Univ.)
Young-Jin Hong (KIRO)
Jun-Soo Jun (NIPA)
Anna Kang (Dankook Univ.)
Dong-Ho Kim (Soongsil Univ.)
Hong-Gi Kim (SNU)
Yoo-Seob Kim (Hallim Univ.)
Joong-Shik Lee (SNU)
Alice Oh (KAIST)
Hyuck-Ro Park (CNU)
Seong-Ki Park (KIST)
Seong-Bae Park (KNU)
Myung-Sook Yeo (Postech)
Sang-Yoon Kim (KUTE)
Seunghoon Choi (Postech)
Min-Seok Kim (KRISS)

Poster Session Co-Chairs
Soo-Yong Shin (Kyung Hee Univ.)
Woon-Taek Woo (KAIST)

Publication Chair
Je-Kwang Ryu (SNU)

Publicity Co-Chairs
Dennis Hong (UCLA)
Seong-Whan Lee (Korea Univ.)
Soo-Hwa Lee (NACS)
Byung-gyu Kim (Sookmyung)
Andrea Thomaz (Georgia Tech)
Kee-Eung Kim (KAIST)
Seokhee Jeon (KHU)

Finance Chair
Sang-Geog Lee (Catholic Univ.)

Local Arrangement Chair
Taikyeong Ted. Jeong (SWU)

Web Chair
Yung-Kyun Noh (SNU)
### PROGRAM

#### Day 1: Thursday, October 27

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:30 - 09:15</td>
<td>Registration</td>
</tr>
<tr>
<td>09:15 - 09:30</td>
<td>Welcome and Introduction</td>
</tr>
<tr>
<td></td>
<td>Byoung-Tak Zhang, Program Co-Chair</td>
</tr>
<tr>
<td></td>
<td>Choong Seon Hong, General Co-Chair &amp; President of KIISE</td>
</tr>
<tr>
<td></td>
<td>Yong-Soo Kim, Ministry of Science, ICT, and Future Planning</td>
</tr>
<tr>
<td>09:30 - 11:30</td>
<td>Instructible Social Robots</td>
</tr>
<tr>
<td></td>
<td>Matthias Scheutz, Director, Human-Robot Interaction Laboratory</td>
</tr>
<tr>
<td></td>
<td>Tufts University</td>
</tr>
<tr>
<td>11:30 - 13:00</td>
<td>Lunch &amp; Poster Session 1</td>
</tr>
<tr>
<td></td>
<td>Chair: Joong-Shik Lee (SNU), Kwang-Soo Cho (Yonsei Univ.)</td>
</tr>
<tr>
<td>13:00 - 15:00</td>
<td>Robot Intelligence vs Biological Intelligence</td>
</tr>
<tr>
<td></td>
<td>Aaron Sloman, Honorary Professor, Artificial Intelligence and Cognitive Science</td>
</tr>
<tr>
<td></td>
<td>University of Birmingham</td>
</tr>
<tr>
<td>15:00 - 15:30</td>
<td>Coffee Break</td>
</tr>
<tr>
<td>15:30 – 17:30</td>
<td>Natural and Artificial Intelligence: Towards Neuromorphic Computational Systems</td>
</tr>
<tr>
<td></td>
<td>Klaus Mainzer, Professor Emeritus, Philosophy and Philosophy of Science</td>
</tr>
<tr>
<td></td>
<td>Technische Universität München</td>
</tr>
<tr>
<td>17:30 – 20:00</td>
<td>Reception &amp; Poster Session 2</td>
</tr>
<tr>
<td></td>
<td>Chair: Joong-Shik Lee (SNU), Kwang-Soo Cho (Yonsei Univ.)</td>
</tr>
</tbody>
</table>

#### Day 2: Friday, October 28

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>09:30 - 11:30</td>
<td>Digital Companions for Seamless Co-Creation</td>
</tr>
<tr>
<td></td>
<td>Andreas Dengel, Scientific Director, Knowledge Management Department</td>
</tr>
<tr>
<td></td>
<td>German Research Center for Artificial Intelligence (DFKI)</td>
</tr>
<tr>
<td>11:30 - 13:00</td>
<td>Lunch &amp; Poster Session 3</td>
</tr>
<tr>
<td></td>
<td>Chair: Taikyeong Jeong (SWU)</td>
</tr>
<tr>
<td>13:00 - 15:00</td>
<td>A Bayesian Perspective of Intelligence</td>
</tr>
<tr>
<td></td>
<td>Hideki Asoh, Senior Research Scientist, Artificial Intelligence Research Center(AIRC), AIST, Japan</td>
</tr>
<tr>
<td>15:00 - 15:30</td>
<td>Coffee Break</td>
</tr>
<tr>
<td>15:30 - 17:30</td>
<td>New Frontiers of Cognitive Systems (Panel Discussion)</td>
</tr>
<tr>
<td></td>
<td>Byoung-Tak Zhang (Moderator), Matthias Scheutz, Aaron Sloman, Klaus Mainzer, Andreas Dengel, Hideki Asoh</td>
</tr>
<tr>
<td>17:30 – 18:00</td>
<td>Award &amp; Farewell</td>
</tr>
</tbody>
</table>
Instructible Social Robots
Matthias Scheutz
Director, Human-Robot Interaction Laboratory
Tufts University

ABSTRACT

There are many ways for social robots to learn new information, from unsupervised data-driven approaches to supervised instruction-based approaches. In this presentation, I will focus on social robots that can interact with humans in natural language and can acquire new knowledge from dialogue-based natural language instructions. I will demonstrate throughout the talk how robots of different sophistication can benefit from natural language instructions and not only quickly acquire new knowledge, but also use it right away during task performance. I will then argue that this way of learning not only mirrors human abilities and can thus also be used to model human instruction-based learning, but moreover allows for unprecedented knowledge acquisition for future robots that can outperform standard statistical approaches.

BIO

Matthias Scheutz is a Professor in Cognitive and Computer Science in the Department of Computer Science at Tufts University. He earned a Ph.D. in Philosophy from the University of Vienna in 1995 and a Joint Ph.D. in Cognitive Science and Computer Science from Indiana University Bloomington in 1999. He has more than 250 peer-reviewed publications in artificial intelligence, natural language processing, cognitive modeling, robotics, and human-robot interaction. His current research focuses on complex cognitive robots with natural language capabilities.
ABSTRACT

Alan Turing died in 1954. The Meta-Morphogenesis project is a conjectured answer to the question: what might Alan Turing have worked on if he had continued several decades after publication of his 1952 paper The Chemical Basis of Morphogenesis, instead of dying two years later?
The project has many strands, including identifying what needs to be explained -- e.g. how could evolution have produced the brains of mathematicians like Pythagoras, Archimedes and Euclid?; or the brains of human toddlers who seem to make and use topological discoveries before they can talk? Or the brains of intelligent non-humans, like squirrels, weaver birds, elephants and dolphins?
How did those ancient human brains make their amazing, deep discoveries over two millennia ago -- long before the development of modern logic or proof-theory?
What features of the “fundamental construction kit” (FCK) provided by physics and chemistry made that possible?
What sorts of “derived construction kits” (DCKs) were required at various stages of evolution of increasingly complex and varied types of biological information processing? Were some currently unrecognized forms of information processing required that will be needed by future Archimedes-like robots -- e.g. in order to be able to discover that extending Euclidean geometry with the neusis construction allows arbitrary angles to be trisected? A major task of the project is collection and analysis of examples of natural intelligence that current AI cannot match, and current neuroscience cannot explain, to help steer research towards new subgoals.
One of my goals is to explain why Immanuel Kant was right about the nature of mathematical discovery in 1781 even if he missed some important details. The presentation will be a revised version of my IJCAI 2016 tutorial. An introduction and some messy notes are here: http://www.cs.bham.ac.uk/research/projects/cogaff/misc/sloman-tut-ijcai-2016.html [Still being revised and extended.]

BIO

Honorary Professor of Artificial Intelligence and Cognitive Science (Retired since 2002, but still doing research full time)
School of Computer Science, The University of Birmingham, UK
Born 1936 Southern Rhodesia (now Zimbabwe).
BSc Mathematics and Physics Cape Town, 1956.
Was seduced from mathematics to philosophy while a graduate student in Oxford, because the philosophers there mis-described mathematics.
Learnt about AI in 1969 from Max Clowes (pioneering vision researcher) at Sussex University. Presented a challenge to AI at IJCAI 1971. Was invited to Edinburgh University to work with Al people on the challenge for a year 1972-3. Spent several more decades working on those problems first at Sussex then Birmingham, linking AI, Philosophy, Cognitive Science, and Biology -- trying to formulate and answer questions ignored by most AI researchers. Found progress very slow, very difficult (and often lonely).
Was inspired by the Turing centenary 2012 to reflect on how Turing might have addressed the problem. This led to the Meta-Morphogenesis project (see my abstract). This has produced a steady flow of (still incomplete) interconnected, freely available, online papers with this root:
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/meta-morphogenesis.html
Natural and Artificial Intelligence:
Towards Neuromorphic Computational Systems
Klaus Mainzer
Professor Emeritus, Philosophy and Philosophy of Science
Technische Universität München

ABSTRACT

In the past, artificial intelligence followed the digital paradigm of computability and the Turing test. But, in natural systems, perception, action and cognition are based on analog abilities which cannot be completely reduced to the digital paradigm. In evolution, natural intelligence has emerged in brains with analog and digital principles. Increase of natural intelligence was realized by increase of the density of more and more neurons in brains with slow synapses, analog weights, and sensible „wetware“ (cellular tissue + neurochemistry), but saving energy. In technology, increase of artificial intelligence was realized by increase of computational velocity and storage of digital computers with “robust hardware” (e.g. silicon + semiconductor technology), but at the cost of high energy.

Mathematically, it can be proven that neural nets (“brains”) and appropriate automata and machines (“computers”) are equivalent – from simple automata and networks up to analog networks with real computing. The lecture considers the computational foundations of analog systems with respect to applications to perception, action and cognition. The target of research should be the convergence of evolutionary (“analog”) and technical (“digital”) strategies in neuromorphic systems which combine technical efficiency with evolutionary advantages (e.g. saving energy).

References:

BIO

Prof. Dr. Klaus Mainzer studied mathematics, physics, and philosophy (1968-1972), Ph.D. (1973) and habilitation (1979) at the university of Münster; Heisenberg-scholarship (1980); 1981-1988 professor for foundations of exact sciences at the University of Constance, vice-president of the university of Constance; 1988-2008 chair for philosophy of science, dean, director of the Institute of Philosophy (1989-2008) and founding director of the Institute of Interdisciplinary Informatics (1998-2008) at the University of Augsburg; since 2008 chair for philosophy of science and technology, director of the Carl von Linde-Academy (2008-2015) and 2012-2014 founding director of the Munich Center for Technology in Society (MCTS) at the Technical University of Munich (TUM); since 2016 TUM Emeritus of Excellence; member of several academies and interdisciplinary organizations (e.g., The Academy of Europe/Academia Europaea in London, European Academy of Sciences and Arts in Salzburg, National Academy of Science and Engineering (acatech) in Berlin) and guest-professor in Brazil, China, Japan, South-Korea, Russia, and USA.

Research interest are mathematization and computer modelling, complex dynamical systems, self-organizing and autonomous systems, complexity and computability, artificial intelligence (AI), computational brain, embodied robotics, and big data technology.

Author of books, e.g.: Thinking in Complexity. The Computational Dynamics of Matter, Mind, and Mankind (Springer 5th enlarged edition 2007, Chinese, Japanese, and Russian translations); The little Book of Time (German: C.H. Beck 5th edition 2005, English and Korean translations); Life as Machine? From Systems Biology to Robotics and Artificial Intelligence (German: Mentis 2010); The Universe as Automaton (with L.O. Chua, Springer 2011); Local Activity Principle. The Cause of Complexity and Symmetry Breaking (with L.O. Chua, Imperial College Press 2013); The Computation of the World. From Big Bang to Big Data (German: C.H. Beck 2014); Artificial Intelligence (German: Springer 2016); Information (Berlin University Press 2016).
Beyond PACS2016
October 27-28, 2016
Beyond AlphaGo

Friday, October 28 / 09:30 - 11:30

Digital Companions for Seamless Co-Creation

Andreas Dengel
Scientific Director, Knowledge Management Department
German Research Center for Artificial Intelligence (DFKI)

ABSTRACT

The momentum of the modern world increasingly requires a rapid and situational learning of new skills. Due to the growing information intensity, the trend towards shorter innovation cycles and the reduction of knowledge half-life time degrades the importance of factual knowledge. In order to enhance performance and productivity, computers are increasingly taking over the role of an amplifying partner supporting our individual handling of diverse information sources and exploring synergies between large communities. In such an evolutionary cyber-social environment, new potentials for digital companions are emerging, assisting users in understanding, learning, decision-making, and memorizing.

This talk discusses the various factors of digitalization and presents examples of current research and development that will affect our way of cognitive co-creation in the near future. It is trying to give some answers to questions, such as: What is expertise and what is the prerequisite for it? What factors influence the creation of competence today? How can technology be employed to act as associative memories for supporting knowledge work in cyber-social settings? How to support knowledge sharing? Can you measure and anticipate information needs? How to employ interactive learning aids for co-creation?

BIO

Professor Andreas Dengel is a member of the Management Board as well as Scientific Director at the German Research Center for Artificial Intelligence (DFKI) in Kaiserslautern where he is leading the Knowledge Management Research Department. In 1993 he became a Professor at the Computer Science Department of the University of. Since 2009 he also holds a Honorary Professorship at the Grad., Computer Science and Intelligent Systems, Graduate School of Engineering of the Osaka Prefecture University. From 1980 to 1986 Andreas studied Computer Science and Economics at the University of Kaiserslautern. He subsequently worked at the Siemens research lab in Munich and at the University of Stuttgart where he completed his doctoral thesis in 1989. In 1991 he worked as a guest researcher at Xerox Parc in Palo Alto. Andreas is a member of various advisory boards, such as for the Cyber-Physical Systems (CPS) IIP Programm of MEXT in Japan, the Computer Vision Center (CVC) at the University of Barcelona, Spain, the Center for Co-Evolutionary Social Systems at Kyushu University, Japan the Center of Excellence on Semantic Technologies at MIMOS in Kuala Lumpur, Malaysia, and the Int’l Conference on Document Analysis and Recognition (ICDAR).

He further is an active reviewer for many organizations, such as the German Council of Science and Humanities, the Research Council of Norway, the Dutch, the Swiss, Luxembourg, and the Austrian Science Foundation. Formerly, he acted e.g. as a member of the IT-Summit Working Group on “Service and Consumer-Oriented Information Technology” consulting the German government on questions of future IT strategies. He also was the German representative in the IST Prize Executive Jury of the European Council of Applied Science (Euro-CASE), a lecturer of the Joint Executive MBA course at the Johannes Gutenberg University at Mainz, the University of Texas at Austin, and the Dongbei University of Finance and Economics at Dalian, China, as well as a member of the METTREC Planning Committee (Metadata/Text Retrieval Conference Committee) of the National Institute for Standards and Technology (NIST) in the United States. Furthermore he was a member of the McKinsey Business Technology Award Committee between 2011 and 2015.

Besides his many keynotes at international conferences, Andreas gave invited talks at well-known universities and research institutions, such as NII, MIT, Stanford University, PARC, UC Berkeley, UC Davis, CMU, ATR, MS Research, Hitachi CRL, Tokyo University, Chinese Academy of Science, and Google Research. Andreas was program/technical chair of international conferences, such as ICPR, ICDAR, ICFHR, DAS, KES, Ki, ICMU and KM. He is an editorial board member of international journals, such as IJDAR, Intelligent Decision Technology, and Future Internet. Moreover, he is founder or initiator of several successful start-up companies. In 2005 he received a “Pioneer Spirit Award” for one of his start-up concepts and at CeBIT 2015 his recent start-up digipen technologies has received the CeBIT Innovation Award. He is co-editor of various international computer science journals, i.e. IJDAR and of book series on Machine Perception and Artificial Intelligence (World Scientific), has written or edited 11 books and is author of more than 340 peer-reviewed scientific publications, several of which received a Best-Paper Award. He supervised more than 200 PhD, master and bachelor theses. In Cambridge, UK, in 2004, Andreas Dengel has been elected a Fellow of the International Association for Pattern Recognition (IAPR). His scientific contributions have been honored several times by international scientific prizes. His main scientific emphasis is in the areas of Smart Data, Deep Learning, Document Understanding, Semantic Technologies, Information Retrieval, Multimedia Mining, and Social Media.
A Bayesian Perspective of Intelligence

Hideki Asoh
Senior Research Scientist,
Artificial Intelligence Research Center (AIRC), AIST, Japan

ABSTRACT

Intelligence has been developed for surviving and behaving well in the changing real world. Observation of the world is inevitably partial, and simulations based on the fundamental principles can’t cover various phenomena in the world. In order to cope with the situation, intelligence seems to use statistical models which are based on probability theory. Using the models, various kinds of information processing such as prediction, simulation, and planning of actions can be done for behaving better in the uncertain real world.

Bayesian inference with probabilistic models is a powerful and mathematically sound tool to treat the uncertainty and have been exploited to construct various intelligent systems such as pattern recognition systems, anomaly detection systems, and so on. There are several evidences supporting that our brain also utilizes Bayesian inference. In my talk, first I would like to give an introduction of Bayesian inference with probabilistic models. In the talk I’l try to show a unified framework in which various concepts in machine learning and probabilistic inference e.g. generative models, discriminative models, predictive coding, deep learning, and so on are related to construct intelligence.

Then, in the latter half of the talk, I would like to introduce the artificial intelligence research center of AIST and show some research activities regarding Bayesian modeling and inference in the context of intelligence embedded in the real world.

BIO

Hideki Asoh received B.Eng. in mathematical engineering and M.Eng. in information engineering from the University of Tokyo, in 1981 and 1983 respectively. In April 1983, he joined in Electrotechnical Laboratory as a researcher. From 1993 to 1994 he stayed at German National Research Center for Information Technology (GMD) as a visiting research scientist. He is currently a deputy director of Artificial Intelligence Research Center (AIRC) in National Institute of Advanced Industrial Science and Technology (AIST). His research interests are in constructing intelligent systems which can learn through interactions with the real world.
<table>
<thead>
<tr>
<th>POSTER INDEX</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>01</strong> Motor Imagery EEG Classification using Sub-Band Regularized Common Spatial Pattern</td>
<td>20</td>
</tr>
<tr>
<td>Sang-Hoon Park and Sang-Goog Lee(The Catholic University of Korea)</td>
<td></td>
</tr>
<tr>
<td><strong>02</strong> Invisible wire between the drone and pilot</td>
<td>22</td>
</tr>
<tr>
<td>Jongwoo Jeon, Sanghyeong Yu and Kwangsu Cho(Yonsei University)</td>
<td></td>
</tr>
<tr>
<td><strong>03</strong> The GRIM: Target Marketing in Social Network</td>
<td>24</td>
</tr>
<tr>
<td>Ashis Talukder, Md. Golam Rabiul Alam, Anupam Kumar Bairagi, Sarder Fakhrul Abedin, Hoang T. Nguyen and Choong Seon Hong(Kyung Hee University)</td>
<td></td>
</tr>
<tr>
<td><strong>04</strong> A New Cognitive System for Drone, Gesture Controlling</td>
<td>26</td>
</tr>
<tr>
<td>Youngjae Yoo, Yunjung Kim and Kwangsu Cho(Yonsei University)</td>
<td></td>
</tr>
<tr>
<td><strong>05</strong> Maximizing Sum-Rate of LTE Users Considering QoS and Co-existence Issue in Unlicensed Band</td>
<td>28</td>
</tr>
<tr>
<td>Anupam Kumar Bairagi, Tran H. Nguyen and Choong Seon Hong(Kyung Hee University)</td>
<td></td>
</tr>
<tr>
<td><strong>06</strong> Emotion Recognition based on Physiological Sensors and Machine Learning Techniques</td>
<td>30</td>
</tr>
<tr>
<td>Anderson Cruz, Gabriel Leitao, Raimundo Barreto(Federal University of Amazonas), Fernando Koch(Korea University) and Tiago Primo(Samsung)</td>
<td></td>
</tr>
<tr>
<td><strong>07</strong> Predicting the Contents popularity with the help of Big Data and In-network Caching in Information Centric Wireless Networking</td>
<td>32</td>
</tr>
<tr>
<td>Kyi Thar and Hong Choong Seon(Kyung Hee University)</td>
<td></td>
</tr>
<tr>
<td><strong>08</strong> ProCCN: Proximity-based Content Centric Networking</td>
<td>34</td>
</tr>
<tr>
<td>Anselme Ndikumana and Choong Seon Hong(Kyung Hee University)</td>
<td></td>
</tr>
<tr>
<td><strong>09</strong> Open Source Base Physical Therapy Games For Parkinson Patient</td>
<td>36</td>
</tr>
<tr>
<td>Safyzan Salim, Mohd Zain Ismail and Farahanis Roslan(Universiti Kuala Lumpur BMI)</td>
<td></td>
</tr>
<tr>
<td><strong>10</strong> Arrhythmia Classification Method Using KNN and Fitness Rules</td>
<td>38</td>
</tr>
<tr>
<td>Woo-Hyuk Jung, David Lee and Sang-Goog Lee(The Catholic University of Korea)</td>
<td></td>
</tr>
<tr>
<td><strong>11</strong> Speeding Up SVM Training in Brain-Computer Interfaces</td>
<td>40</td>
</tr>
<tr>
<td>David Lee, Woo-Hyuk Jung and Sang-Goog Lee(The Catholic University of Korea)</td>
<td></td>
</tr>
<tr>
<td>POSTER INDEX</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td></td>
</tr>
<tr>
<td><strong>12</strong></td>
<td>An Analysis on Variability in Motor Learning of Patients with Parkinson's Disease through a Virtual Throwing Task</td>
</tr>
<tr>
<td></td>
<td>Jinseok Oh, Je-Kwang Ryu(Seoul National University), Se-Woong Park(Northeastern University) and Kyoung-Min Lee(Seoul National University Hospital)</td>
</tr>
<tr>
<td><strong>13</strong></td>
<td>Visual Imagination from Texts</td>
</tr>
<tr>
<td></td>
<td>Hanock Kwak and Byoung-Tak Zhang(Seoul National University)</td>
</tr>
<tr>
<td><strong>14</strong></td>
<td>KAP Based PICO-Compliant Query Construction</td>
</tr>
<tr>
<td></td>
<td>Muhammad Afzal and Sungyoung Lee(Kyung Hee University)</td>
</tr>
<tr>
<td><strong>15</strong></td>
<td>Model-Driven Adaptive User Interface based on Context and User Experience Evaluation</td>
</tr>
<tr>
<td></td>
<td>Jamil Hussain and Sungyoung Lee(Kyung Hee University)</td>
</tr>
<tr>
<td><strong>16</strong></td>
<td>IoT based Human Centric Context Awareness Framework for Healthcare and Wellness Platform</td>
</tr>
<tr>
<td></td>
<td>Jaehun Bang, Dong Uk Kang, Tae Ho Hur, Thien Hyun The, Muhammad Asif Razzaq, Wajahat Ali Khan(Kyung Hee University), Oresti Banos(University of Twente) and Sungyoung Lee(Kyung Hee University)</td>
</tr>
<tr>
<td><strong>17</strong></td>
<td>Service Curation Framework for Context-Aware Personalized Recommendations</td>
</tr>
<tr>
<td></td>
<td>Ali Syed Imran, Muhammad Sadiq, Muhammad Afzal and Sungyoung Lee(Kyung Hee University)</td>
</tr>
<tr>
<td><strong>18</strong></td>
<td>Smart CDSS: Knowledge Acquisition and Modeling using Mind Maps and Decision Trees</td>
</tr>
<tr>
<td></td>
<td>Musarrat Hussain, Maqbool Hussain and Sungyoung Lee(Kyung Hee University)</td>
</tr>
<tr>
<td><strong>19</strong></td>
<td>Semantic Reconciliation Model for interoperable and shareable knowledge authoring environment</td>
</tr>
<tr>
<td></td>
<td>Taqdir Ali and Sungyoung Lee(Kyung Hee University)</td>
</tr>
<tr>
<td><strong>20</strong></td>
<td>UAV-assisted Intelligent Crowdsourcing in Natural Calamity</td>
</tr>
<tr>
<td></td>
<td>Sarder Fakhrul Abedin, Md Golam Rabiul Alam, Anupam Kumar Bairagi, Ashis Talukder and Choong Seon Hong(Kyung Hee University)</td>
</tr>
<tr>
<td><strong>21</strong></td>
<td>Data Curation Layer for Wellness Platforms</td>
</tr>
<tr>
<td></td>
<td>Hafiz Syed Muhammad Bilal and Sungyoung Lee(Kyung Hee University)</td>
</tr>
<tr>
<td><strong>22</strong></td>
<td>Smartphone User Authentication Based on Rotation Touch Behavior Biometrics</td>
</tr>
<tr>
<td></td>
<td>M.H. Mohd Zain, Toshiyuki Kinoshita(Tokyo University Technology) and M.Z. Ismail(University Kuala Lumpur-BMI)</td>
</tr>
<tr>
<td>Poster Number</td>
<td>Title</td>
</tr>
<tr>
<td>---------------</td>
<td>------------------------------------------------------------------------</td>
</tr>
<tr>
<td>23</td>
<td>Local and Global map interface design for flying drone</td>
</tr>
<tr>
<td>24</td>
<td>Mining Minds, an opensource initiative towards health and wellness platforms</td>
</tr>
<tr>
<td>25</td>
<td>SaKEM: A Semi-automatic Knowledge Engineering Methodology for Building Rulebased Knowledgebase</td>
</tr>
<tr>
<td>26</td>
<td>X-UDeKAM: An Intelligent Method for Acquiring Declarative Structured Knowledge using Chatterbot</td>
</tr>
<tr>
<td>27</td>
<td>Aupair: A Home Robot for Personal Care</td>
</tr>
<tr>
<td>28</td>
<td>Schedulebot: A Home Robot Learning and Acting Schedule Adaptively via Dynamic Environments</td>
</tr>
<tr>
<td>29</td>
<td>Evaluation of User Interface Design using Psychophysiological Measures</td>
</tr>
<tr>
<td>30</td>
<td>Cambot: A Visual Conversation Robot for Interactive Engagement</td>
</tr>
<tr>
<td>31</td>
<td>Glassbot: Personalized Wearable Agents Learning from Everyday Human Behaviors</td>
</tr>
<tr>
<td>32</td>
<td>Jibobot: A Personal Assistant Robot with Social Motions</td>
</tr>
<tr>
<td>Post No.</td>
<td>Title</td>
</tr>
<tr>
<td>---------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>33</td>
<td>Pandabot: Multimodal Story Learning with Dynamic Memory Construction</td>
</tr>
<tr>
<td>34</td>
<td>Knowledgebot: Neuroknowledge based complimentary Learning Model for Question Answering Systems</td>
</tr>
<tr>
<td>35</td>
<td>Cafebot - A Conversational Cashier Robot in Cafes</td>
</tr>
<tr>
<td>36</td>
<td>Pororobot: Child Tutoring Robot for English Education</td>
</tr>
<tr>
<td>37</td>
<td>Storybot: Story Learning from Cartoon Videos via Consecutive Event Embedding</td>
</tr>
<tr>
<td>38</td>
<td>Anomaly Pattern for Fraud Detection and Clustering on Large-scale On-line Enterprise Market</td>
</tr>
<tr>
<td>39</td>
<td>Childbot: A Conversational Assistant for Child Care</td>
</tr>
<tr>
<td>40</td>
<td>Human-machine Interactive Knowledge Discovery - integrating IBM Watson Concept Insight in HCI and future research methodology</td>
</tr>
</tbody>
</table>
Early Registration Deadline: October 25, 2016

**Registration Fee**

<table>
<thead>
<tr>
<th></th>
<th>Early Registration (by Oct. 25, 2016)</th>
<th>After the Early Registration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student</td>
<td>KRW 250,000 / USD 250</td>
<td>KRW 300,000 / USD 300</td>
</tr>
<tr>
<td>General</td>
<td>KRW 500,000 / USD 500</td>
<td>KRW 600,000 / USD 600</td>
</tr>
</tbody>
</table>

**Method of Payment**

* Card
  Payment by credit card is available through the online procedure.

* For Bank Transfer
  Bank Name: WOORI BANK PANGBAEYOK
  BR. SWIFT Code: HVBKKRSEXXX
  Account No.: 463-04-103170
  Account Name: Korea Institute of Information Scientists and Engineers(KIISE)
  Bank Address: 1585, Sangam-dong, Mapo-gu, Seoul, Korea

**Registration Policy**

- Cancellation should be made by Oct. 25, 2016 for 100% refund. It is not refundable after Oct. 25, 2016.
- Please check the Registration list
- Contact: Heesoo Choi, 02-588-9246 / hschoi@kiise.or.kr

**On-line Registration:**
http://www.kiise.or.kr/pacs/2016/
The symposium will be held at Grand Hall of the AT Center on the 5th floor (address: 27, Gangnam-daero, Seocho-gu, Seoul, Korea; Google Maps).

The AT center is in the outskirt of Seoul, the capital city of South Korea, and is located 6 miles south of the downtown Seoul and easily reachable by subway and bus.

aT Center, Seoul, Korea

Subway Station: Yangjae Citizen’s Forest(Maeheon), Shinbundang-Line

http://atcenter.at.or.kr/contents/acen331100/view.action
Location: 5th floor

Grand Conference Room 5F
POSTER SESSION
Motor Imagery EEG Classification using Sub-Band Regularized Common Spatial Pattern

Sans-Hoon PARK
Department of Digital Media Engineering
The Catholic University of Korea
Bucheon
timesea821@naver.com

Sang-Goog LEE
Department of Media Technology & Media Contents
The Catholic University of Korea
Bucheon
sg.lee@catholic.ac.kr

Abstract—Common Spatial Pattern (CSP) algorithm shows limited performance in Small-Sample Setting (SSS) situations because they rely on covariance matrix. In addition, they show large differences in performance depending on frequency bands being used. To solve these problems, 4-40Hz band EEG signals are divided into nine sub-bands and Regularized CSP (R-CSP) is applied to individual sub-bands. Fishers Linear Discriminant (FLD) is applied to the features of R-CSP extracted from individual sub-bands and the results obtained through the foregoing are connected for all sub-bands to make an FLD score vector. Thereafter, Principal Component Analysis (PCA) is applied to use the FLD score vectors as inputs of the classifier Least Square Support Vector Machine (LS-SVM). The proposed method particularly excellent performance in SSS situations.

Keywords—brain-computer interface (BCI); common spatial pattern (CSP); electroencephalogram (EEG); motor imagery; classification;

I. INTRODUCTION

The Brain-Computer Interface (BCI) technology is a means of communicate with the outside world without using the peripheral nervous system by converting humans brain activities into control commands [1]. The Common Spatial Pattern (CSP) which is a spatial filtering technique was proposed for effective classification of multi-channel electroencephalogram (EEG) signals. The CSP is quite effective in classifying motor imagery EEG [2]. However, the CSP has several shortcomings. First, the classical CSP method greatly relies on sample based covariance matrix estimation. Therefore, the performance is limited when the number of training samples is small [3]. However, the CSP has several shortcomings. First, the classical CSP method greatly relies on sample based covariance matrix estimation. Therefore, wide frequency bands are set or frequency bands that fit individuals features are manually selected in general. In this paper, a method is proposed for simultaneously overcoming the covariance matrix estimation problem in small sample setting (SSS) situations and dependency on frequency bands. The performance of proposed method is evaluated using BCI competition III dataset IVa and indicated as classification accuracy for five subjects.

II. METHODOLOGY

A block diagram of the method proposed in this paper is shown in Figure 1.

A. Regularized Common Spatial Pattern (R-CSP)

Applying regularization technologies to the CSP was proved as being effective in solving the high bias and variance of covariance matrix estimation that may occur in SSS situations [5]. The computation of RCSP covariance follows the following procedure. The sum of the sample covariance matrix of $M$ training experiments in class $i$ is as follows:

$$C_i = \sum_{m=1}^{M} C_{(i,m)}$$

In the regularization process, regularization term $\hat{C}_i$ is created using EEG collected from other experimental subjects.

$$\hat{C}_i = \sum_{m \neq i}^{M} C_{(i,m)}$$

where $\hat{\eta}$ is the total number of class $i$ experiments obtained from other experimental subjects. $\hat{C}_i$ reduce variance and creates reliable results in covariance matrix estimations. Covariance matrix $\hat{\beta}$ that includes not only the covariance matrix of a certain experimental subject but also the covariance matrices of other experimental subjects is defined as follows:

$$\hat{\beta} = \left(1 - \beta\right) \cdot C_i + \beta \cdot \hat{C}_i$$

Finally, the regularized average covariance matrix of the R-CSP of individual classes is as follows:

$$\tilde{\Sigma}_i(\beta, \gamma) = (1 - \gamma) \cdot \hat{\beta} + \gamma \cdot \text{tr} [\hat{\beta}] \cdot 1$$

where $\mathbf{I}$ is the $N \times N$ identity matrix, $\beta$ and $\gamma$ are regularization parameters ($0 \leq \beta, \gamma \leq 1$). $\beta$ controls the variance of covariance matrix estimations and $\gamma$ plays the role of reducing large eigenvalues and increasing small eigenvalues to...
obstruct the occurrence of bias. When \( \beta = \gamma = 0 \), the R-CSP is the same as the CSP. Hereinafter, the method of extracting the features of the R-CSP is the same as the classical CSP method. In this study, the R-CSP is applied to individual sub-bands to use two spatial filters from the first rows and the last rows respectively, \( m = 2 \).

### B. Fisher’s Linear Discriminant

The FLD is an algorithm to find the axis necessary to achieve the maximization of between-class scatter and the minimization of within-class scatter. The problem of calculating the FLD projection matrix can be solved by finding the projection matrix \( W_{FLD} \) that derives the maximum values of Fisher’s object functions

\[
W_{FLD} = \arg \max_{W} W^T S_{B} W = S_{B}^{-1} (m_1 - m_2)
\]

### C. Principal Component Analysis

The PCA is a useful method for feature extraction and reduction of the dimension of feature vectors. The purpose of the PCA is to find the linear orthogonal transformation matrix that maximally maintains the variance of features.

### III. DATA DESCRIPTION

BCI Competition III Dataset IVa was used to measure the performance of the algorithm proposed in this paper [6]. This dataset is publicly available and useful for evaluation of the performance of algorithms using small numbers of training signals. EEG data were recorded using 118 channels in five healthy subjects (aa, al, av, aw, and ay). In this study, the 100Hz version was used and out of the motor imagery EEG data of each experimental subject for 3.5 sec., those from 0.5 to 2.5 sec. were used. In addition, as shown in Figure 2, a total of 18 channels were used.

![Fig. 2. Locations of electrodes used](image-url)

### IV. RESULT AND DISCUSSION

The performance of the SBRCSP algorithm is verified by comparing it with the CSP, R-CSP, FBBCSP, SBCSP and FBRCSP methods. In the results of the proposed method, the numbers in the brackets are the parameter sets obtained through five-fold cross validation. As shown in Table I, the SBRCSP method improves mean classification accuracy by 14.99%, 9.55%, 8.63%, 6.95% and 4.18% compared to the CSP, R-CSP, FBBCSP, SBCSP and FBRCSP. In the classification results, attention should be particularly paid to experimental subject ay. The number of training data of ay is 28 which is the smallest compared to other experimental subjects. The performance for ay was improved by 22.62% compared to the CSP method. This result indicates that the SBRCSP method is quite powerful for processing SSS situations. The SBCSP method may unstable covariance matrix estimations in SSS situations because it relies on CSP covariance matrix estimations. However, the SBRCSP method shows better performance in SSS situations because it uses R-CSP covariance matrix estimation method.

### V. CONCLUSION

In this paper, a method was proposed that can simultaneously solve the covariance matrix estimation problem in SSS situations and the problem of dependency on frequency bands. No study has been conducted to overcome these two problems simultaneously thus far. The results of the study based on 18 channels of BCI competition III dataset IVa verified that the SBRCSP method had good effects on processing of motor imagery EEG signals. In particular, the SBRCSP method was more effective in SSS situations. Also, the methods that divided EEG signals into sub-bands were shown to be not very effective in SSS situations but quite effective in large training set while the R-CSP was shown to have large effects in SSS situations. In conclusion, the method in this paper suggested a guideline to overcome the dependence of frequency range and limited performance in SSS situation.

### REFERENCES


### TABLE I. CLASSIFICATION ACCURACY OF BCI COMPETITION III DATASET IVa

<table>
<thead>
<tr>
<th>Subject</th>
<th>CSP</th>
<th>R-CSP</th>
<th>FBBCSP</th>
<th>SBCSP</th>
<th>FBRCSP(5, 7)</th>
<th>SBRCSP(5, 7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>aa</td>
<td>85.17</td>
<td>85.17</td>
<td>85.17</td>
<td>85.17</td>
<td>85.17</td>
<td>85.17</td>
</tr>
<tr>
<td>al</td>
<td>85.17</td>
<td>85.17</td>
<td>85.17</td>
<td>85.17</td>
<td>85.17</td>
<td>85.17</td>
</tr>
<tr>
<td>av</td>
<td>85.17</td>
<td>85.17</td>
<td>85.17</td>
<td>85.17</td>
<td>85.17</td>
<td>85.17</td>
</tr>
<tr>
<td>aw</td>
<td>85.17</td>
<td>85.17</td>
<td>85.17</td>
<td>85.17</td>
<td>85.17</td>
<td>85.17</td>
</tr>
<tr>
<td>ay</td>
<td>85.17</td>
<td>85.17</td>
<td>85.17</td>
<td>85.17</td>
<td>85.17</td>
<td>85.17</td>
</tr>
<tr>
<td>mean</td>
<td>85.17</td>
<td>85.17</td>
<td>85.17</td>
<td>85.17</td>
<td>85.17</td>
<td>85.17</td>
</tr>
</tbody>
</table>
Invisible wire between the drone and pilot
Evaluation of single-handed remote controller

Jongwoo Jeon  
Graduate Program in Cognitive Science, Yonsei University  
Seoul, Republic of Korea  
mashi79513@gmail.com

Sanhyeong Yu  
Graduate Program in Cognitive Science, Yonsei University  
Seoul, Republic of Korea  
shyu.ux@gmail.com

Kwangsu Cho  
Graduate School of Information; Graduate Program in Cognitive Science, Yonsei University  
Seoul, Republic of Korea  
kwangsu.cho@gmail.com

Abstract—To fly drone safely and easily, we have to solve the misalignment problem in external piloting. There are two kinds of misalignment which occurred under additional cognitive operating. First, misalignment can be arisen when pilot change forward/backward or right/left direction after yaw function which rotating the drone. In addition, as finger movement of pilot and drone operation are misaligned, it is difficult to use controller for pilot embody. Therefore, we develop and study a single-handed remote controller which can solve the inappropriate yaw function problem and finger movement and drone operation misalignment problem of two-handed controller.

Keywords—unmanned aerial vehicles(UAVs); single-handed controller; embodiment;

I. INTRODUCTION

As civil drones or unmanned aerial vehicles(UAVs) have come into widespread use in diverse areas, drone accidents are also increasing by novice pilots who never had expert drone piloting training. Out of 32 percent of all relevant drone accidents were caused by a pilots’ mistakes while controlling the drone [1]. Drones accidents damaging human and property will be increasing without safety secure. Therefore, a solution which let the novice pilots fly drone safely and easily is necessary.

Methods of controlling drone are critical issue causing drone accidents [2, 3]. There are two main methods of controlling drone: internal piloting and external piloting. Internal piloting refers to the controlling method through first person view as if pilot sitting inside drone. Internal piloting has been widely used in military drones through built-in camera view or global positioning system(GPS) map. On the other hand, external piloting refers to the controlling method through third person view. Pilots see the drone with their own eyes and use two-handed remote controller which has been used for decades.

External piloting is likely to cause more accidents because of the conflict between pilots’ perspective and drones’ actual operational perspective [2, 3, 4, 5] under the external piloting circumstance, pilots should convert their own visual field perspective to the drones’ actual operational perspective. The conflict and cost arisen from the converting process make pilots react slowly, thus increase the risk of accidents. This conflict is also called as misalignment. The Federal Aviation Administration(FAA) recently mitigated regulations for civil drones under the external piloting circumstance [6]. Consequently, pilots will be exposed to accidents more frequently.

However, the drone controller which has been used in general for civil drones is difficult to be used under the external piloting circumstance. In the past military drones, internal piloting was the main drone controlling method so that the controller was specialized in internal piloting circumstance. Although existing controller is not proper to use for external piloting, the controller is generally used with low cost civil drones. Thus, pilots cannot avoid the misalignment: they see the drone, but they have to ignore it and think as if sitting inside drone.

There are two possible reasons in detail. First, the existing drone controller has an inappropriate function for the external piloting. The yaw function rotates the drone and it is appropriate function to internal piloting. However, as the reference axis of the forward/backward(pitch) and right/left(roll) functions, which are on the right side, is changed by the outcomes of the yaw function then the misalignment is arisen under external piloting.

The other reason is that the controller has a difficulty to be embodied because the mapping between finger movement and drone operation is misaligned. For example, when pilot try to move drone toward the upward, pilot has to move throttle joystick toward the forward. The controller has two separated dimensions to implement three-dimension of actual drone operation space: one-dimension on the left hand(up/down) side and two-dimension(right/left/forward/backward) on the right hand side. It is an effective way to distribute the work loads of drone control into two hands. However, in the point of embodiment view, a gap is generated between the dimension of the actual drone operation space and the separated dimensions of the two-handed controller.

In this research, we develop and study a single-handed remote controller which can solve the inappropriate yaw function problem and finger movement and drone operation misalignment problem of two-handed controller. The single-
handed remote controller and the drone operates as if the drone is connected to the head of the controller by an invisible wire. The yaw function, which makes a lot of misalignments, is removed in this solution, thus pilots can always control drone with their own perspective. Also, single-handed controller integrated the separated dimension of two-handed controller into the one side then pilots can hold controller with one hand. The controller is composed with two parts: sensing and patterning of pilot’s arm movement, and only forward/backward(throttle) joystick.

II. EVALUATION

This solution has two main development parts. First, it is necessary to classify the movement pattern of pilot’s arm and hand into up/down/right/left direction. Second, the moving distance of drone should follow the outcomes of trigonometrical function calculated by the rotating distance of controller and the distance between drone and controller.

Then we will evaluate our solution with two experiments. We need to find out the effects of Fitts’ law first [7, 8]. Fitts’ law is a kind of evaluation method of interface measuring a travelling time from start point to target. The outcomes of Fitts’ law tasks, accuracy and movement time, can be applied to drone pilot task in three dimensional space. Second, path planning task would be implemented to find out how pilot can fly a drone safely and easily.

III. CONCLUSION

The two-handed controller for drones has been used for decades. It is stable using controller with two hands. However, it is hard to use easily under external piloting circumstance because of yaw function and inappropriate embodiment.

This research proposes a single-handed controller to avoid yaw function misalignment and embodiment through integrated dimensions into one hand. The solution can make novice pilots fly drone freely and easily with the effects of embodiment of the controller.

ACKNOWLEDGMENT

This research was supported by Yonsei IoT-UX Academy, a training business for creative academy(CAC15113) funded by Seoul government and Seoul Business Agency.

This material is based upon work supported by the Ministry of Trade, Industry & Energy(MOTIE, Korea) under Industrial Technology Innovation Program. No.10060517, ’Development of an user-centered product design support system based on cognitive and affective information’

REFERENCES


The GRIM: Target Marketing in Social Network

Ashis Talukder, Md. Golam Rabbiul Alam, Anupam Kumar Bairagi, Sarder Fakrul Abedin, Hoang T. Nguyen and Choong Seon Hong

Department of Computer Engineering, Kyung Hee University, South Korea
Email: {ashis, robi, anupam, saab0015, nguyenth, cshong}@khu.ac.kr

Abstract—Influence maximization (IM) is a popular research area for viral marketing in social network. In this research we propose a novel problem, Reverse Influence Maximization (RIM) problem for target marketing along with a greedy solution GRIM model under Linear Threshold (LT) model. The GRIM model determines the opportunity cost returned by minimum number of nodes that must be activated in order to motivate a target seed set. We also perform simulation to evaluate the performance of the algorithm using two real world datasets.

I. INTRODUCTION

The Influence Maximization (IM) for viral marketing has gained remarkable interest in social network research. The IM calculates a fixed size seed set that can activate the maximum number of nodes in the network [1]. Here we propose a novel problem for target marketing, named Reverse Influence Maximization (RIM) problem. It estimates the opportunity cost [2] which is the minimum number of nodes that must be activated in order to motivate a given seed set. These targeted seed members are supposed to be prominent entity. For example, Usain Bolt’s Facebook photo using a Sumsung cellphone can influence his millions of fans to buy the same phone. Thus the GRIM model also have great business value like the IM.

In this research we prove that the RIM problem is NP-Hard and propose a Knapsack based greedy solution under classical LT model [1]. We also evaluate the performance of the GRIM model with two real social network datasets.

The rest of the paper provides literature review, problem formulation, GRIM model, performance evaluation and conclusion in the consecutive sections.

II. LITERATURE REVIEW

To the best of our knowledge the IM problem was first introduced by [3] in 2001. Then Kempe et al. in [1] give a great shape with two classical models such as Linear threshold (LT) and Independent cascade (IC) models providing \((1-\frac{1}{e})\) performance approximation.

In [4], the authors have proposed a heuristic approximation for outbreak detection using IM. Goyal et al.[5] extend it which has 35 – 55% faster running time. Chen et al. introduce a degree discount (DC) heuristic in [6]. It improves the accuracy of [1] and the running time of [4] simultaneously. Most recently, the authors in [7] have given a running time improvement by an innovative idea of stop and stare sampling.

But none of the above has addressed the problem of finding the opportunity cost for target marketing that we propose.

III. PROBLEM FORMULATION

Assume a social network is given by a graph \(G=(V,E)\), where each vertex is a users and each edge is a social relationship. A user \(u\) influences \(v\) with weight \(w_{uv}\). The node \(v\) is activated if the influence coming from all the active in-neighbors is no less than a given threshold \(\theta_v\) that is \(\sum_{v\in \Gamma^{\setminus}(u)} w_{uv} \geq \theta_v\) [1]. Here \(x_u\) indicates whether a node \(u\) is active \((x_u = 1)\) as input layer node or not \((x_u = 0)\) and same definition holds for \(h_u\) and \(x_u\) for hidden and target layer nodes respectively.

For a given seed set \(S\) of \(K\) influential customers, the RIM aims at finding opportunity cost set denoted by \(\Gamma(S)\) and the opportunity cost \(\sigma(S) = |\Gamma(S)|\).

![Diagram of Influence Maximization](image)

Figure 1: The RIM reduce the calculation level by a neural network analogy.

The RIM decomposes the network into \(K\) neural networks containing \(v\) as the only one output layer node as depicted in the Fig. 1. The marginal opportunity cost \(\sigma(v)\) for each target node \(v\) is computed by:

\[
\sigma(v) : \text{minimize} \sum_{u \in \Gamma^{\setminus}(v)} x_u
\]

subject to,

\[
\sum_{u \in \Gamma^{\setminus}(v)} w_{uv} x_u \geq \theta_v,
\]

\[
x_u \in \{0, 1\}, w_u \in \{0, 1\}
\]

Then the final opportunity cost and is given by:

\[
\sigma(S) = \bigcup_{v \in S} \sigma(v)
\]

IV. SOLUTION FRAMEWORK OF RIM

We propose the GRIM model which is Knapsack based greedy solution of the RIM problem.
A. Meeting the Challenges

We set the limit of calculation with a simple 3-layer neural network analogy as shown in the Fig. 1. The decomposition results in three cases. The target node \( v \) has zero (A), one (B) or multiple layer (C) of in-neighbors. The Case A is trivial and we set \( \sigma(v) = |\Gamma(v)| = |\{v\}| = 1 \). The Case B is the basic unit of calculation and the Case C is the combination of multiple A and B cases. We set \( \theta_v \) to some smaller value to avoid insufficient influence. A node \( u \) may be activated as input layer node of a target node \( v_1 \) and also activated as hidden layer and/or a target layer node of another target node \( v_2 \). Then the node \( u \) will not be added in the opportunity cost set and this optimization is called commonality discount.

**Theorem 1.** The RIM problem is NP-Hard.

**Proof.** The RIM can be reduced to Knapsack problem by considering node’s threshold as Knapsack size, influence weights as item weights and substituting the objective function of the RIM problem by a maximize format of the negation of the equation (1).

Algorithm 1: GRIM Model

```
Input: \( G(V,E), S \)
Result: \( \sigma(S), \Gamma(S) \)

1 \( \Gamma(S) = \emptyset \); /\* Hidden to output layer */
2 for each \( v \in S \) do
3 \( \Gamma(S) = \Gamma(S) \cup \Gamma(v) \) by equation (1) to (3);
4 end
5 \( S_1 = \Gamma(S), \Gamma(S) = \emptyset \); /\* Input to hidden layer */
6 for each \( v \in S_1 \) do
7 \( \Gamma(S_1) = \Gamma(S_1) \cup \Gamma(v) \) by equation (1) to (3);
8 end
9 \( \Gamma(S) = \Gamma(S_1) \); /\* Commonality discount */
10 for each \( v \in \Gamma(S) \) do
11 \( \Gamma(S) = \Gamma(S) - \{v|h_v = 1 OR t_v = 1\} \);
12 end
13 return \( \sigma(S) = |\Gamma(S)| \);
```

B. The GRIM algorithm

The GRIM algorithm is stated in the Alg. 1. The complexity of the algorithm is given in equation (5) where \( D \) = maximum number of neighbors (degree) in the network.

\[
\mathcal{T} = O(K \cdot D^2)
\]

V. PERFORMANCE EVALUATION

We evaluate the performance of GRIM model for two real datasets. They are scaled and stated in the Table I.

<table>
<thead>
<tr>
<th>Networks</th>
<th>ego-Facebook [9]</th>
<th>Epinions [9]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nodes</td>
<td>1000</td>
<td>1000</td>
</tr>
<tr>
<td>Edges</td>
<td>15,969</td>
<td>40,240</td>
</tr>
</tbody>
</table>

The GRIM model exhibits the expected trend of cost increasing with the increase of threshold when the seed size is constant as portrayed in the Fig. 2 for both the datasets. It also shows good running time in the Fig 3 since it involves fixed number of layers and hence lesser nodes and edges.

VI. CONCLUSION

In this research we propose the Reverse Influence Maximization (RIM) problem along with greedy solution GRIM model to estimate the opportunity cost for target marketing in social network. The GRIM model show good running time (order of \( K \cdot D^2 \)) with reasonable opportunity cost.

The main limitation of this research is that the GRIM model can not always provide optimal solution for its greedy nature but yet provides a feasible solution.

REFERENCES


A New Cognitive System for Drone, Gesture Controlling*

Youngjae Yoo, Yunjung Kim and Kwangsu Cho
Graduate Program of Cognitive Science
Yonsei University
Seoul, Korea

Abstract—The number of accidents of drone controlling is increased as personal drones being popular. We can say that people who are not professional for controlling drone are exposed to the serious danger of the accident. To make controlling drones more intuitive and easier, pre-programming methods which let drones fly as planned or gesture - controlling methods were suggested. In this study, natural gesture design for controlling drones more intuitive is suggested. By taking a metaphor from joystick which users can face commonly when controlling the motion of an object in the digital interface, immediacy and naturalness were increased. In the case of controlling drones, users have to go outside which is a too variable environment in the brightness of light or distance between drone and user. For this reason, it is better to use Data – Glove method than Vision Based method to implement gesture controller for drone since Vision Based method is critical to angle and brightness of light. Using suggested gesture design which came from joystick metaphor, users will be able to control drones with four basic commands more intuitive and easier without the long period of training.

Keywords—unmanned aerial vehicle; gesture control; cognitive system;

I. INTRODUCTION
The number of accidents of drone controlling is increased as personal drones being popular. Considering that 33% of drone accidents in U.S Army were caused by pilots, we can say that people who are not professional for controlling drones are exposed to serious danger of accident [1]. As a result, regulations for drones are being discussed actively in many countries and even some countries decided to make strict laws for drone in order to satisfy their own purpose. For this reason, there are various studies on going about controlling drones more intuitive and easier to mitigate difficulties in dealing with drones.

To make controlling drones more intuitive and easier, pre-programming methods which let drones fly as planned or gesture - controlling methods were suggested [2]. As we interact with all kinds of objects by hands in everyday life, gestures are very natural and intuitive tools for interaction especially [3]. So gesture interface will help users control drones easier and safer.

In this study, natural gesture design for controlling drones more intuitive will be suggested. This study so that, will be able to provide safer environment to drone users who are entry – level while controlling drones by preventing exposure to danger of accidents. When developing gesture controller, there are some problems like “Midas Touch”, gesture recognition accuracy, or stolen by stranger. These problems have to be considered to develop effective controller using these gestures.

II. RELATED WORK
It is adequate to use hand – gesture design among some kinds of gesture design methods to control drones. Since hands are the most frequent body part interacting to other objects [3]. Motion gesture can be divided into two kinds, one is gesture mapping and the other is physical characteristics [4]. Gesture mapping is about connection between user motion gesture and command to device. Physical characteristic is about effect of motion gesture to kinetic stimulus or complexity of gesture. To make gesture mapping smooth, using metaphor which refers to something as being the same as another thing is needed. On the other hand, to satisfy the physical characteristic classifying drone control command has to be done first. There are four commands in common quadrotor drone which has four wings. The first one is “Throttle” which controls flying up and down. And the next one is “Yaw” which controls rotating clock wise and counter clock wise. The third one is “Pitch” which controls moving forward and backward. And the last one is “Roll” which controls moving left and right [5]. Designed gestures which are for replacing four commands Throttle, Yaw, Pitch and Roll should be simple through adjusting kinetic stimulus and dimensions.

III. SOLUTION
Joystick was a metaphor of the gestures designed in this study. By taking metaphor from joystick which users can face commonly when they control motion of object in digital interface, immediacy and naturalness were increased. Controlling with gestures like using virtual joystick is not only natural but also advantageous to improve gesture recognition or task performance [6]. To start the gesture control, users have to clench their fist like grabbing joystick as preparation. At this position, Pitch and Roll command can be easily operated by tilting the fist forward/backward and left/right. In case of
Throttle, double - tilting the fist forward quickly means going down. And lastly in case of Yaw, similar as Throttle, double – tilting the fist left or right side means left and right rotation respectively.

There are two ways to implement suggested gesture design, “Data – Glove based” method and “Vision Based” method. The Data – Glove method uses motion of hands and fingers as a parameter by digitizing through sensors. Vision Based method, on the other hand, can recognize gestures through analyzing images only by camera excepting any other extra devices [7]. In case of controlling drones, users have to go outside which is too variable environment in brightness of light or distance between drone and user. For this reason, it is better to use Data – Glove method than Vision Based method to implement gesture controller for drones since Vision Based method is critical to angle and brightness of light.

IV. CONCLUSION

Using suggested gesture design which came from joystick metaphor, users will be able to control drones with four basic commands more intuitive and easier without long period of training. Through this study of gesture design, users can control drones with simple gesture controller instead of heavy controllers as a drone. This will increase usefulness of drone controlling and reduce the accidents by controlling drones mistakenly as well.

ACKNOWLEDGMENT

The This research was supported by Yonsei IoT-UX Academy, a training business for creative academy(CAC15113) funded by Seoul government and Seoul Business Agency.

This material is based upon work supported by the Ministry of Trade, Industry & Energy(MOTIE, Korea) under Industrial Technology Innovation Program. No.10060517, ‘Development of an user-centered product design support system based on cognitive and affective information’

REFERENCES

Maximizing Sum-Rate of LTE Users Considering QoS and Co-existence Issue in Unlicensed Band

Anupam Kumar Bairagi
Dept. of Computer Engineering
Kyung Hee University
South Korea
Email: anupam@khu.ac.kr

Nguyen H. Tran
Dept. of Computer Engineering
Kyung Hee University
South Korea
Email: nguyenth@khu.ac.kr

Choong Seon Hong
Dept. of Computer Engineering
Kyung Hee University
South Korea
Email: cshong@khu.ac.kr

Abstract—Both academia and industry are engaging different new technologies to deal with the current traffic cyclone in cellular networks. Long Term Evaluation-Advanced (LTE-A) is such a recent technology that serves heavy mobile traffic. But with insufficient licensed spectrum, LTE-A also cannot meet the quality of service (QoS) requirements of all its users. So, by augmenting the benefits of LTE-A into unlicensed spectrum known as LTE-U, we can boost the performance of 4G/5G cellular network. The process will sure to conflict with other technologies which are already using same unlicensed band. Moreover, if multiple cellular network operators (CNOs) use the same unlicensed band then they will diminish the benefits of each others. In this paper, we explore the CA of licensed and unlicensed spectrum when QoS of user cannot be met with licensed spectrum by deploying dual mode small cell base station (SBS) and considering minimum requirement of unlicensed WiFi access points (WAP). Here, we try to solve this problem with the help of Nash bargaining game (NBG) between LTE-U and WAP by cooperative approach.

I. INTRODUCTION

During the past decade, mobile data traffic has already shown exponential outgrowth and in the next five years, it is anticipated to flourish the amount 1000 times[1]. To meet this huge demands, new technologies like LTE or LTE-A, massive multiple-input multiple-output (MIMO), device-to-device (D2D) communication are coming forward with scarce licensed spectrum. CNOs are trying to meet this huge users’ demand by deploying SBSs which require low-cost and low power with the help of using technique of licensed spectrum.

But this actions are not enough to cope with the exponentially increasing data traffic and meet the stringent QoS of emerging wireless services in the next generation cellular system using limited licensed spectrum. So, some CNOs have already deployed WAPs to offload part of their traffic in unlicensed band. But such initiatives are not so effective due to the inferior performance of WiFi technology and not cost effective as CNOs also need to invest on backhaul and core network to integrate WAPs with cellular system. This shortcomings can be overcome by extending the blessings of LTE-A in the unlicensed spectrum known as LTE-U. This can be technically assured via the use of CA technology which was standardized in LTE Releases 10-12. LTE-U is already inaugurated (part of the LTE Release 13) to allow consumers for accommodating licensed and unlicensed carrier under a single LTE network infrastructure[2].

As LTE-U and WiFi are diminishing each others performance and within themselves, this interactions can be modeled as a game theory framework namely bargaining game to improve their performance. There are several proposals to coexist fairly of this two, but few of them have considered inter-operator interaction and nobody has seen it in the eye of bargaining game. In this paper, we have tried to maximize LTE-U sum-rate considering the QoS requirement of the users and co-existence issue with WAPs by using bargaining game. Here, we formulate the problem as an optimization one and then solve it using bargaining game.

II. SYSTEM MODEL AND PROBLEM FORMULATION

As SCNs are feasible solution to meet data demand of the users, cellular operators are deploying more and more SBSs to facilitate growing services. This ultra dense nature of SBSs from different operators bound to conflict with each other and also with local WAPs if they want to operate in unlicensed spectrum to provide guaranteed QoS. As each operator can control the interference between MBS and its SBSs, we are considering an environment where there is a set of dual-mode LTE-A SBSs, $S = \{1, 2, ..., N\}$ operated by $N$ different operators and a set $W = \{1, 2, ..., M\}$ of $M$ WAPs. Each SBS $i$ can serve...
downlink operation of maximum users $U_i$ at a time with it’s licensed spectrum $B_i$. Both SBSs and it’s associated users are distributed randomly in the area of interest. As only one user can be served by WAP at a time, we assume that there are $M$ WiFi users distributed randomly in the same area. Both SBSs and WAPs operate in the same unlicensed band $B_u$. SBSs work in SDL mode with CSAT and CA technology.

When LTE-U user $j$ of SBS $i$ is using both spectrum then the achieved rate of that user by using Shannon’s capacity shown in equation (1).

$$R_j^i = (b_{ij} + b_{u,ij}) \log_2(1 + \frac{(b_{ij} + b_{u,ij})P_j h_{ij}}{\sum_{k \in S, k \neq i} b_{kj}^k P_k h_{kj} + P_w h_{w,j} + \sigma^2})$$

But study [3] shows that WiFi presence affects negligibly to the LTE-U performance. So we can ignore the interference generated by WiFi system to LTE-U user and equation (1) reforms like shown in equation (2).

$$R_j^i = (b_{ij} + b_{u,ij}) \log_2(1 + \frac{(b_{ij} + b_{u,ij})P_j h_{ij}}{\sigma^2})$$

From equation (2), we find that a major part of interference is like NBG which provides a unique solution of the problem. In that case there is no interference from other SBSs and equation (2) changes it’s form like:

$$R_j^i = (b_{ij} + b_{u,ij}) \log_2(1 + \frac{(b_{ij} + b_{u,ij})P_j h_{ij}}{\sigma^2})$$

According to study [4], the saturation capacity of $M$ WiFi (APs employ CSMA/CA with binary slotted exponential back-off) users sharing same unlicensed bandwidth is shown in the equation (3).

$$R^W = \frac{P_i P_j E[P] M^{−1}}{(1 - P_i T_c) T_c + P_i P_j T_c + P_i (1 - P_j) T_c}$$

$R^W$ is achievable when WiFi network only access the unlicensed channel. But if WAPs and SBSs are deployed in the same conflicting area, then WiFi users get almost no access in the channel and achieve an insignificant data rate. So, for fair coexistence of WiFi and LTE-U need to share the time slot in such a way that WAPs can maintain a minimum data rate and SBS can guarantee the QoS of its users. As LTE-U system manages the physical resource in a centralized manner rather than DCF of WAPs, SBSs need to decide appropriate portion of time to achieve minimum rate of each WAP. So, when SBSs give $T \in [0, 1]$ time slot to WAPs then the achievable rate of LTE-U user and WAP are shown in the equations (6) and (7) respectively.

$$R_j^i(T) = (b_{ij} + (1-T)b_{u,ij}) \log_2(1 + \frac{(b_{ij} + (1-T)b_{u,ij})P_j h_{ij}}{\sigma^2})$$

$$R^W(T) = R^W T$$

III. SOLUTION WITH NBG

Bargaining game is a typical cooperative game that is fair in case of resource allocation. So, this interaction can be considered as a two player bargaining game shown in Figure 2, where $P = \{L, W\}$ are the set of players. Let $S$ be a closed and convex subset of $R^{|P|}$ which represents the set of feasible payoff allocations that the players can achieve if they cooperate using utility function $U_i(T_i), \forall i$. Let $d$ be a set of disagreement payoff. Then the ordered pair $(S,d)$ is called a $|P|$-player bargaining game.

**Theorem 1:** There exists a unique solution concept $\phi(S, d)$ that satisfies all six axioms of Nash bargaining solution and it follows (7) where $B = \{ (r_1, r_2) \in S | r_2 \geq Q_w \}$[5].

$$r^* = \phi(S, d) \in \arg\max_{x \in B} \prod_{i=1}^{|P|} (r_i - d_i)$$

**Theorem 2:** The point in which just the minimum rate of WAPs are satisfied is the solution of the bargaining game.

IV. CONCLUSION

In this short paper, we have tried to meet the QoS requirements of the users by augmenting unlicensed spectrum with licensed one in LTE-A network. Here, LTE-A network use SDL scheme to take advantages of unlicensed spectrum while maintaining minimum requirement of other WAPs who use the same unlicensed band in the conflicting region. For this we have solved the problem by utilizing cooperative approach like NBG which provides a unique solution of the problem.

REFERENCES

Emotion Recognition based on Physiological Sensors and Machine Learning Techniques
Anderson Cruz, Gabriel Leitão, Raimundo Barreto, Tiago Primo and Fernando Koch

Abstract—The main aim of this paper is to present a method of collection, analysis and emotion recognition in digital education context, using physiological sensors (Galvanic Skin Response and Heartbeat) and machine learning techniques. The education environments can provide many possibilities around the human variations what is promising to experimentation. The approach used to stimulate the emotions was the use of several different videos. We evaluated which machine learning technique has the best performance to recognize the following emotions: anguish, anxiety, anger, sadness and joy. The adopted performance metric was F-Score. Experimental results show that the kNN and C4.5 techniques achieved highest accuracy in recognizing emotions.

Index Terms—Emotional Recognition and Perception, Physiological Signals, Digital Education

I. INTRODUCTION

Because of the several possible environments and situations, the education domain has a great variability of human activities and behaviors and, therefore, human signals related to inertial or emotion. Thus, the problem to be dealt with in this paper is to recognize emotions through video sessions. In the digital education context, the emotion recognition can be a great ally in the generation of personalized teaching methods. For instance, the data collected from heartbeat rate and skin conductance – combined with other information from content interaction [1] – allow to recognize emotions like joy, anguish, anxiety, neutrality, anger, sadness and others. We propose to combine this extra information about user behavior with methods for recommendation systems and adaptive education, allowing to uninterrupted monitoring and adjustment of educational activities. In this paper, we introduce the proposal, the developed methods, and the results from practical experiments.

II. PROPOSED METHOD

We propose a method to collect physiological signals – i.e. skin conductance and heartbeat rate – together with videos to stimulate various emotions. The wearable sensors can provide information to train methods of supervised machine learning. The objective is recognize students’ behavior from the data collected and analyzed. In addition, we also apply methods to noise reduction, feature extraction and normalization.

The method works as depicted in Figure 1. The first phase of the proposed method is collecting data from GSR and heartbeat through emotion stimulus. The next phase is

![Fig. 1. Proposed Method](image-url)

Fig. 1. Proposed Method

the pre-processing where noise removal and characteristics extraction are carried out. The third phase is the data analysis where several statistics are generated to understand the data. The classification phase adopts supervised machine learning techniques to generate a model that is able to classify instances by inferring a function from labeled training data (input object and a desired output value). Such inferred function can be used for mapping new samples. In this case, the inferred function model is able to distinguish emotions from the labeled training data. Finally, the Evaluation Model estimates the performance of the classification process through the F-Score metric. In this case, we compared several machine learning techniques to find the one with best performance.

III. EXPERIMENT

Figure 2 presents our experimentation platform. It includes a Samsung Gear S device – for data storage and communication...
Changes in emotional state reflect in physiological signals that are captured and processed by our model. For instance, the emotion “anxiety” involve increase of heartbeat rate, shortness of breath, dizziness, intermittent hot and cold, and others [3]. We are able to capture few nuances through the sensors and infer the possibility with a level of accuracy.

We executed a controlled experiment applying the proposed platform upon 43 students in 8 videos lasting about 1 hour. The videos were selected among those that stimulate emotions and cause changes in the physiological signals. After the video session all participants filled a form basically including which kind of emotion each participant has felt in specific times.

From the GSR, HR sensors and the form, we collected raw data like timestamp, ID Student, ID Video, GSR Value, BH Value, ID Session, ID Sensor, age and gender of student. After a process of removing of noises and computation of information gain, we have obtained one set with the following features: ID Student, GSR Value, BH Value and Emotion. This dataset has 51,099 instances and it was used as input to the supervised machine learning techniques, where the system learns how to classify the emotions based on feature sets provided during the training phase [4]. To avoid overfitting, it was necessary divide the data in two groups (Tables I and II), according to amount of instances per emotion.

We adopted the metric F-Score to evaluate the classification algorithm performance. This method allows to quantify the balance between the total emotions recognized when compared to the total number of emotions that should be correctly recognized, based on the correlation:

$$F - Score = \frac{2 \cdot \text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$  \hspace{1cm} (1)

<table>
<thead>
<tr>
<th>TABLE I: F-score separated by emotion, using a window size of 1 second to classify: neutrality, joy and others.</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Correct classification by emotions</td>
</tr>
<tr>
<td>-------------------------------------</td>
</tr>
<tr>
<td>Neutrality</td>
</tr>
<tr>
<td>Joy</td>
</tr>
<tr>
<td>Others</td>
</tr>
</tbody>
</table>

We have used the software WEKA for data processing and inference of the emotions according to each chosen machine learning method [5].

Table I presents the accuracy of the correct classification of the neutrality, joy and others for the all algorithms tested. The term 'others' refers to emotions not present in the classification adopted. Experimental results shows that kNN (with k=7 neighbors) and C4.5 have obtained the higher accuracy in the classification of these three emotions.

Table II presents a second set of results from the other group of emotions (anger, anxiety, anger and sadness). In this case, we observed that, except the anger, the most of the emotions has accuracy near to 80%. We believe that the “anger” emotion was lower due to video presented had many variability in emotions. The emotion with highest skin conductance was anxiety and with the greatest heartbeat was sadness.

IV. CONCLUSION

This paper presented a method to emotion recognition in education context. We collected and analyzed data from Galvanic Skin Response and Heartbeat Rate sensors and we compared some machine learning methods to recognize the following emotions: joy, neutrality, sadness, anxiety, angry and anguish in order to evaluate the accuracy of each one.

Because of the difference in the quantity of instances per emotion, we divide the emotions in two groups. In both groups, kNN and C4.5 achieved the highest accuracy. In the first group, kNN achieved 77.50% of average accuracy and C4.5 obtained 77.46%. In the second, 84.1% and 83.8%, respectively.

As future work we propose to change the implementation to a less intrusive method based on wearable computing devices. Thus, the sensors can be embedded in the clothing, allowing for continuous monitoring, collecting information about the motion, in addition to physiological signals. Additionally, we need to improve the accuracy of the sensors and recognition methods, and how to optimize the power consumption.

ACKNOWLEDGMENT

This paper was sponsored by Samsung Eletrônica da Amazônia Ltda under Brazilian Federal Law No. 8248/91, and FAPEAM through project 582/2014 (PROTI-PESQUISA).

REFERENCES

Predicting the Contents popularity with the help of Big Data and In-network Caching in Information Centric Wireless Networking

Kyi Thar, Choong Seon Hong
Department of Computer Science and Engineering, Kyung Hee University, 446-701, Republic of Korea.
Email: {kyithar, cshong}@khu.ac.kr

Abstract—Currently, the mobile data traffic is tremendously increasing and the solution is needed to handle this increasing demands. There are several solutions to solve this issue. Among them, the Information-Centric Networking (ICN) is one of the most promising solutions, where ICN reduces the network traffic by caching the contents temporarily. So, ICN-enabled Base Station (BS) can provide these cached contents to the user, instead of retrieving from the original server. Also, ICN-enabled BSs aggregate the same content requests from different users at the Base Stations, instead of forwarding every request to the original server. Thus, the decision to cache the contents is important and it should be effectively reduced the data traffic. Even tough, there are several caching decision algorithms already existed, we can still improve the cache decision for large scale network with the help of big data (the information stored by the operators). Therefore, in this paper, we proposed a caching scheme which is working together with the big data platform to predict the popular contents and cache these contents at the BSs efficiently. We used Collaborative Filtering (CF) to fill up the sparse matrix with prediction values. Then we predict the popular contents list by using the Machine Learning (ML) technique with the consideration of others context information. Depending on the recommend list, BSs store the contents and send back the feedback to the controller to update the loss function.

Index Terms—Content Centric Networking, Big Data, In-Network Caching, Collaborative Filtering.

I. INTRODUCTION

The mobile data traffic growth is exponentially increasing, especially on video traffic. In order to reduce network data traffic, the Information-Centric Networking (ICN) [1] is proposed. The basic concept of the ICN is that the routers or nodes store the contents temporarily and provide these contents to the users, instead of retrieving from the original server. Among the ICN architecture, the Content Centric Networking [2] architecture is popular. Therefore, for the 5G network, we employed the ICN concept at the Radio Access Network, where Base Stations (BS) and Small Cell Base Station (SBS) are attached with cache space. Therefore, ICN-enabled BSs and SBSs can store the contents temporarily and provide these contents to their users, instead of retrieving from the original server. In order to improve the performance of the ICN caching, we used the benefit of Big Data technology [3] [4] to predict the contents popularity, generate the recommended list to cache the contents for each BS and SBS.

Previously, the contents popularity is assumed as the Zipf distribution. Depending on the Zipf distribution, proposed a different kind of cache decision algorithm. For the current work, we consider predicting the content popularity and cache the contents depending on the popularity values by analyzing a large amount of user information. The challenges of this work are as follows. Matrix completion problem is NP-hard. It is difficult to solve overfitting and cold start problem, and popularity prediction mechanism with contextual information such as (event, the new movie or not).

II. SYSTEM MODEL

The network model is shown in Fig. 1, where the content servers are located at the outside of the current Autonomous System (AS) and the controller is located at the Data Center. The complex calculation for predicting the content popularity and recommending the contents list to a store of each BS are done at the controller. The BSs and SBSs are located inside the RAN, where BSs and SBSs have attached with the cache space to temporary store the contents. All BSs and SBSs are connected with the core network/controller via the backhaul link. For the simplicity, we define the BSs and SBSs of the InP as BSj, where j = 1, ..., J. There are total number of file F,
where $F = 1, 2, \ldots, f$. The cache capacity of the BS$_j$ is denoted as $Z_j \in Z$, where $Z$ is the total cache capacity of the whole network in Terabytes. The BS$_j$ cannot store the contents more than that it cache capacity, $\sum_{f=1}^{F} x_{jf} \leq Z_j, \forall j, f$, where $x_{jf}$ is the binary variable and $s_f$ is the size of the content $f$. The BS$_j$ stores the content $f$ when $x_{jf} = 1$, otherwise $x_{jf} = 0$. The users, those are connected with the BS$_j$, are denoted as $u_j \in U, a_{u_j}$ is the user association indicator, where $a_{u_j} = 1$ means that user $u_j$ associates to BS$_j$, otherwise $a_{u_j} = 0$. Each user can only associates to only one BS, which is defined with $\sum_{j=1}^{J} a_{u_j} = 1, \forall u$.

The overview process of our proposed scheme is shown in Fig.2, which consists of five steps. 1) The first step is the information collecting stage or recommendation requesting stage. Every time $t$ BSs and SBSs send their local information to the controller to get the recommendation list. By using these recommendation lists, BSs and SBSs make the cache decision. 2) In the second step, the controller constructs the sparse matrix. We used Collaborative filtering(CF) to estimate the missing values. 3) In the third step, we minimize the least square error of estimating the missing values by using the Alternating Least Square (ALS) method. Then, we get the non-sparse matrix with predicted/estimated values. 4) In the fourth step, controller filters and constructed the augmented popularity matrix with the contents popularity values from the non-sparse matrix. 5) In the fifth step, the controller generates an ordered recommendation list to store the contents for each BS and SBS.

III. ESTIMATE THE MISSING VALUES

To estimate the missing values, we use CF, where CF systems can be categorized into two groups: memory-based and model-based methods. Memory-based methods simply memorize the rating matrix and issue recommendations based on the relationship between the queried user and item and the rest of the rating matrix. Model-based methods fit a parameterized model to the given rating matrix and then issue recommendations based on the fitted model. Memory-based methods store the rating matrix and estimate based on the relationship between user and item. Model-based methods construct the model with training data. Then predict the unseen ratings with that constructed model.

The base station-content matrix is denoted as $P = \{p_{ij}\}_{n \times m}$, where $p_{ij}$ is the popularity value of the content $i$ at the BS$_j$, $n_b$ is the number of BSs and $n_m$ is the number of contents. Then, we estimate the missing values of $p_{ij}$ in $P$ with a low-rank approximation of the base station-content matrix $P$. This approach models both BSs and contents in a low dimensional feature space. Each BS and each content have a feature vector. Each popularity of a content of BS is a low dimensional feature space. Each BS and each content has a feature vector. Each popularity of a content of BS is denoted as $f_{nxm}$, where $p_{ij} = b_j^T c_i$. Then we define the to minimize the root mean square error,

$$\min_{b, c} \sum_{i,j,k} (p_{ij} - \hat{p}_{ij})^2 + \lambda (\|b_i\|^2 + \|c_j\|^2).$$

(1)

To solve the eq(1), we used ALS as in [5].

IV. POPULARITY PREDICTING

The controller predicts the content popularity depending on the context as in [6]. We consider, item fatigue, tiredness and temporal features, similarity or diversity features. From this step, we can get the ordered content popularity list. Then send the lists to each base station.

V. CONCLUSION

In this paper, we proposed a cache decision process with the help of Big Data, where do we use the Collaborative Filtering to estimate the missing popularity values and then we predict the content popularity by using context information. To estimate the missing popularity values, we applied Alternative Least Square and minimize the root square error. For the future work, we will run and test our proposed scheme at the SPARK environment.

REFERENCES

ProCCN: Proximity-based Content Centric Networking

Anselme Ndikumana
Department of Computer Science and Engineering
Kyung Hee University, Rep. of Korea
Email: anselme@khu.ac.kr

Choong Seon Hong*
Department of Computer Science and Engineering
Kyung Hee University, Rep. of Korea
Email: cshong@khu.ac.kr

Abstract—With the increase of global mobile and smart device users, such as smartphone, smart watch, tablet, etc. motivates the end-users to create and share the contents within proximity location or on Internet. The existing proximity based services such as mobile social network, marketing, etc. exploit the location of user equipment for service discovery and information sharing, where service discovery and service information is mapped with end-user equipment location. However, proximity discovery may violate the end-user privacy through sharing location information. The Content Centric Networking (CCN) proposed as an alternative to the current IP based networking, where the content is requested by name rather than the IP address, can help in sharing content without sharing end-user location. In this paper, we proposed proximity-based Content Centric Networking (ProCCN), which is related to proximity content and service discovery without using location information. End user request called Interest Packet can be used to retrieve the content and service information.

I. INTRODUCTION

In the recent couples of decades, the number of contents, mobile & smart devices, and means to communicate over internet has grown rapidly, and in this heterogeneity of these entities, Information Centric Networking, a new Internet architecture Content Centric Networking (CCN) has been introduced, in which considers content as primitive [1].

As of now, there is no several studies related to CCN mobility [2], but with the increase of global mobile and smart device users such as smartphone, smart watch, tablet, etc., CCN offers more opportunity for mobile end-users. In CCN Contents can be shared without using host identifier or location information; there is no need for mobile users to update location information during the handoff. On content not received, consumer can resubmit Interest in new location [2]. In view of the above, through the use of CCN, users can communicate among themselves without sharing their updated location information within proximity location or on Internet.

The existing proximity based services such as mobile social network, marketing, etc. exploit the location of user equipment for service discovery, where service discovery and service information is mapped with user equipment location [3]. However, mapping user equipment location and service may violate the end-user privacy who want to keep his location information private. The Content Centric Networking (CCN) proposed as an alternative to the current IP based networking, where the content is requested by name rather than the IP address, can help in sharing content without sharing end-user location.

Fig. 1. Proximity-based Content Centric Networking

In this paper, we proposed proximity-based Content Centric Networking (ProCCN) in which is related in proximity content and service discovery without using location information, but in case the end-user/service provider wants to share his location, he can include his location information in the content.

II. PROXIMITY BASED SERVICE OVERVIEW

One of the key enable technology of Proximity based service is LTE direct in 3GPPP release 13 which require direct communication between device to device (D2D), which offers more advantages such as: (1) user may get high data rates with lower end to end delay due to the short range data communication. (2) The end-user equipment may communicate directly without need cell towers (eNodeB) to connect, especially when cellular coverage fails or is not available. (3) Offloads cellular traffic [4].

LTE direct has a range up to 500 meters, which is far more than WIFI and Bluetooth, and it does not utilize location information in determining proximity. LTE direct uses radio
signals called expressions, which are common languages for applications, and facilitate the applications to discover each other [5].

III. PROXIMITY-BASED CONTENT-CENTRIC NETWORKING (ProCCN)

With the increase of mobile and smart devices usage, the users are not only the consumers of the contents, they participate in contents creation and distribution. This makes mobile and smart devices around us to become the source of contents and services, which can be shared among the users within proximity location, without using public network such as Internet which sometimes requires more delay and payment.

CCN considers content name as central point of communication, where content is requested and retrieve by name. Retrieving content by name can help the end-users in service discovery and content sharing within proximity location without using location information. In our system model described in figure 1, we proposed proximity-based Content Centric Networking, which is related to the proximity content and service discovery without using location information through the use of LTE direct interface. UEs 1, 2, and 3 can retrieve content or service information from Content Provider and Service Provider, while UE 4 can retrieved cached content and service information from UE3.

To get the content through the use of LTE direct interface, end-user requests the content or service information through flooding content request called Interest Packet. Any UE in LTE direct coverage, which has the content or service information, returns the content or service information (with service location) to the requester.

Fig. 2. Content Store table(CS) for End-user Equipment

Normally, the end-user creates the content and stores it in his mobile device. To differentiate the content that the end-user does not want to share, and the CCN content which can be served to any UE on demand; we propose the modification of the Content Store table(CS) or content memory, where the content received from other node(s) will be denoted R, while the content created by end-user will be denoted P. The end-user equipment serves only the content denoted P. However, when the end-users wants to share cached content in R category, he can update content source field by replacing R by P (Ref. Figure 2).

With the help of three main CCN Data structures, namely: (a) Content Store (CS) or memory, (b) Pending Interest Table (PIT), which records unsatisfied Interest packets with their incoming faces, and (c) Forwarding Information Base (FIB) which records outgoing faces, we present our CCN Proximity-based forwarding flowchart in figure 3.

Fig. 3. CCN Proximity-based forwarding

In our proposal, we consider also the mobility of source of content/service or the requester. When source of content/service or the requester change the location, on content not received with Interest life time, end-user/ requester can re-express again Interest in new location.

IV. CONCLUSION

In this paper, we proposed proximity-based Content Centric Networking (ProCCN), which is related to proximity content and service discovery without using the end-user location information through the use of LTE direct. In our proposal, the Interest Packet can be used to retrieve the content and service information. In the future, we aim to extend our work with simulation and more analysis.

REFERENCES


Open Source Based Physical Therapy Games For Parkinson Patient

Safyzan Salim  
Mohd Zain Ismail  
Farahanis Roslan
Communication Technology Section  
Communication Technology Section  
Communication Technology Section
Universiti Kuala Lumpur BMI  
Universiti Kuala Lumpur BMI  
Universiti Kuala Lumpur BMI
Gombak Selangor Malaysia  
Gombak, Selangor, Malaysia  
Gombak, Selangor, Malaysia
safyzan@unikl.edu.my  
mzain@unikl.edu.my  
roslanfarahanis@yahoo.com

Abstract—For the information Parkinson is a progressive disease of the nerves system marked by tremor, muscular rigidity, and slow, imprecise movement, chiefly affecting middle-aged and elderly people. Although the disease is commonly treated by drugs, in recent years there have been growing interest in non-drug management[1], [2]. The allied health professions thus have a major role to play in the rehabilitation of these patients both at home and in hospital[2]. Constant training is a key factor for this type of therapy. Digital games can be one way to enhance. It is very popular form of entertainment for many people[3]. Tapping into this trend, a study will be made to adapt. So, it can be applied to make the emerging field of low-cost computerized physical therapy games,[3] This therapy physical games were designed to improve coordination in people with Parkinson’s disease, a chronic, progressive neuromuscular disease characterized by shaking, slowness of movement, limb and trunk rigidity[1], [2]. It is also focused on specific body movements and gestures[4].

Keywords—Parkinson, digital games, progressive neuromuscular disease, portable

I. INTRODUCTION

In Malaysia the Parkinson patient is increased by year to year[4]. The majority of them aged in range sixty years old and above[1], [4]. Parkinson is a disease of the nervous system[1]. It generally causes the muscles to become stiff and the body to tremble which gradually gets worse, as a person gets older[1], [4]. However, for those who have the incurable condition and have to struggle on with its daily effects in their bodies the important thing to do is to keep fighting back[4]. Most of them always get the treatment in the hospital and listen the advice from the doctor[1], [4]. The various health services available in Malaysia for people with Parkinson are physiotherapy, home visits, medications, and brain surgery[2]. Therefore, they need a treatment that can rehab them without feeling the pain and low cost physical therapy. Moreover it becomes dull and boringness therapy[2]. The best way is, purpose the interesting and fun therapy such as open source based physical therapy games[2], [3]. This physical therapy games can be more effective because exercising through games is perceived as play rather than therapy by user,[5] Other than that, games offer the possibility of unlimited repetitions facilitated by a continuous feedback loop, immersion, and a state of winning[2], [5]. While mainstream commercial training games and game devices like the three sixty degree movement games by using android appliances could be useful for physical therapy purposes for Parkinson patients to help to decrease rigidity. It is also can improve their coordination, spatial perception, reflexes and balance. Furthermore the commercial games such as Playstation, EyeToy, usually include negative feedback when the performance of the player is not sufficient and lack meaningful task in terms of therapy. In contrast to these systems, this physical therapy is particularly design for Parkinson patient[5].

II. METHODS

The Open Source Based Physical Therapy Games more towards the Parkinson patients will play such as rolling ball games. It will be divided into three stages from easy, medium, and harder. The games will start from the easy stage which means the player will roll the ball form the top and roll the ball down by the accelerometer movement and put the ball into the basket. At the medium and harder stages, it is still the same as easy stage to roll the ball from the top to down but they have obstacles. The player will lose if they fail to face these obstacles. I think, it is very suitable with the Parkinson patient because can improve their coordination, spatial perception, reflexes and balance.

A. Methodology

The Open Source Based Physical Therapy Games more towards the Parkinson patients. The patient will play the games by using accelerometer movement. An accelerometer is a sensor which measures the tilting motion and orientation of a mobile phone.

B. Materials

This project consists of several materials that are very important in order to complete the system process. Also to produce the product the core element needs to be specific in term of the instruments used and the capability of each part. There are divided into two requirement which is hardware and software.

1) Hardware Requirement: The hardware requirement consists of two main items. The first item is tablet where the application of the games is stored. The tablet shall be android Operating System (OS) and has 7.0 inches of screen that
suitable for the game applications. The second item is high-end server which is functioning to store the software and design of the games.

2) Software Requirement: Basic4android (B4a) Rapid is application development tool for native Android application. The B4a is applied in this project rather than the alternative to programming such as Java, Android SDK, Visual Basic and Visual Basic.Net

C. Block Diagram

Figure 1 shows the block diagram of the project. It begins with the Parkinson patients. They act as the main problem in this project. After that, they must have the android devices such as the tablet or hand phone. Nowadays the android devices become cheaper. Everybody afford to buy it. The function of the android devices is to install the digital games application. They can install the application by free. Then the Parkinson patient can play the games. By using the android devices, they can play the games either at home or everywhere. So by playing the digital games they can practice the exercise and improve their hand skill. If they go to the hospital, the doctors can take and record their progress or improvement.

III. RESULTS

The Open Source Based Physical Therapy Games consists of two stages as described below:

A. Easy Stage

The ball will roll from the top and to down by tilting motion and put the ball into the basket as shown in Figure 2.

B. Medium Stage

The ball will roll from the top and to down by tilting motion and put the ball into the basket but will have the obstacles as shown in Figure 3. If the ball hit the obstacles the player will lose.

C. Results Analysis

Figure 4 shows an example of the result by using the B4a android software. From the result analysis, it shows that the result of Parkinson patient is more effective as exercising through games is perceived as play rather than therapy. Other than that, training games by using android devices could be useful for Parkinson patient to help them decrease rigidity.

IV. CONCLUSION

The goal of the project is to help the Parkinson patient easy to make their physiotherapy. The Open Source Based Physical Therapy Games For Parkinson Patient is very user friendly to the Parkinson patient. They can use this application in their android devices and save their time to do the therapy at the hospital. The source also have good advantages that can help them recover the disease without feel the pain. Additionally, the Parkinson patient will enjoy to do the exercise games.

The proposed game also are able to improve the movement and motion symptoms of the Parkinson patient which is the most important aspects of a Parkinson’s disease patient rehabilitation. It is also able to improve of the cognitive symptoms like attention span and depression. The proposed game also is a noble alternative method which is a non-drug management for patient with Parkinson.

REFERENCES


Arrhythmia Classification Method Using KNN and Fitness Rules

Woo-Hyuk Jung
Department of Digital Media, Media Engineering
The catholic university of Korea
Bucheon, Gyeonggi, Republic of Korea

David Lee, Sang-Goog Lee
Department of Media Technology & Media contents
The catholic university of Korea
Bucheon, Gyeonggi, Republic of Korea

Abstract—The present study combined features extracted through principal component analysis and linear discriminant analysis and it proposed a method of classifying arrhythmia using the fitness rules and K-nearest neighbor. In the preprocessing phase, the amplitude was adjusted through normalization and baseline fluctuations, and diverse noises existing in signals were removed using wavelets. In the feature extraction and selection phase, the pre-processed electrocardiogram signals were decomposed, and d4 and a4 feature coefficients were extracted. Thereafter, the dimensions of the d4 and a4 feature coefficients were reduced using PCA and LDA, respectively, and the resultant coefficients from the reduction were combined and composed as features. In the classification phase, the test data were classified into four classes using K-nearest neighbor. The goodness of fit of the classified data was tested using the fitness rules. As a result, sensitivity=93.61%, specificity=98.52%, and positive predictive value=91.90% were exhibited.

Keywords—Electrocardiogram; Classification; Arrhythmia

I. INTRODUCTION

ECG represents the electrical activity of the heart, and it exhibits representative signals of cardiac physiology that are usefully used in disease diagnosis. Because ECG signals provide information on heart conditions, studies that accurately analyze ECG signals have been conducted on, e.g., H/W and S/W systems for monitoring [1] and peak detection [2]. For the accurate diagnosis and treatment of patients, the accurate recognition and classification of diverse types of arrhythmia are essential. Inan et al. [3] proposed an algorithm to classify the combined features into normal and arrhythmic beats using the mixture of experts. Hu et al. [4] designed an algorithm for the efficient classification of arrhythmia using the mixture of experts. Khadra et al. [5] proposed a bispectral analysis technique for the classification of four types of arrhythmia. In the present study, an arrhythmia classification method is proposed that uses PCA and LDA feature combination based KNN and the fitness rules. The accuracy of arrhythmia classification is improved by combining the features of PCA and LDA to reinforce the representative characteristics of the original signal and reclassifying those data that have not been accurately classified into classes using the KNN and the fitness rules.

II. EXPERIMENTAL METHOD

The experimental method used in the present study is shown in Fig. 1.

Fig. 1. Block diagram of the suggested method

A. Pre-processing method

To eliminate diverse noises in ECG signals, first, the amplitudes of signals are adjusted through normalization. Thereafter, Daubechies 2 is used to remove baseline fluctuations and power line noises. The original signal is decomposed into seven levels using Daubechies 2. The high frequency component of the lowest level 7 (H7) is set to 0, and an inverse wavelet is performed with the level 7 low frequency component (L7) to create the components of CA6 (inverse wavelet of L7, H7). The inverse wavelets are repeatedly performed up to the upper level 2 to create the components of CA1. Thereafter, the CA1 components are removed from level 1 low frequency components (L1). Finally, signal noises are removed by setting high frequency components (H1) to 0 and performing an inverse wavelet with L1 less CA1.

B. Feature extraction and selection method

In the present study, to include sufficient important data from ECG signals, 128 samples are extracted from the left side of the R peak, 127 samples are extracted from the right side of the R peak, and 256 samples in total are set as a segment. PCA and LDA
are algorithms to extract features and can remove redundant data to reduce dimensions and extract necessary signals. Combining features extracted through PCA and LDA to construct a new feature can help improve the accuracy of classification. Among the reduced features, PC1-PC17 and LD1-LD17 are combined to set feature data containing at least 98% of the original features. The combined feature data are constructed as shown in Eq. (1).

\[
\text{Feature}_{\text{pca, lda}} = \begin{pmatrix}
PC_{11} & LD_{11} \\
PC_{21} & LD_{22} \\
... & ... \\
PC_{17} & LD_{17}
\end{pmatrix}
\] (1)

C. Arrhythmia classification method

The 17 feature PCs and 17 LDs extracted from the test data are set to a two-dimensional matrix and applied to the KNN based on nine neighborhood sizes. The Euclidean distances in each row are calculated and aligned in ascending order so that there are nine neighborhoods. The nine values aligned in ascending order are assigned to their relevant classes, and the class of the largest number of neighborhoods out of the nine aligned neighborhoods is determined as the class of the row. The classes of all 17 rows are determined as such, and the class that obtains the largest number of votes among the selected 17 classes is selected. Thereafter, whether the classified data fit the class is tested using the fitness rules. The fitness rules are set based on the RR interval. In the case of single intervals, there are difficulties in setting clear boundaries between beats [6]. Therefore, in the present study, diverse RR interval rules are set as shown in Tables 1 and 2. If the class of the classified beats satisfies the rules set forth in Table 1 and Table 2, the beats are determined as correctly classified beats; if not, the beats are determined as beats to be reclassified.

TABLE 1. N, S, V, F common rule

<table>
<thead>
<tr>
<th>Rule</th>
<th>Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rule 1</td>
<td>(min RR1 &lt; RR2 &lt; max RR2) &amp; (min RR1 &lt; RR2 &lt; max RR2)</td>
</tr>
<tr>
<td>Rule 2</td>
<td>(median RR1/median RR2 = 0.1) &lt; RR1/RR2 &lt; (median RR1/median RR2 = 0.1)</td>
</tr>
<tr>
<td>Rule 3</td>
<td>(min RR1 &lt; RR2 &lt; range RR1) &amp; (range RR1 &lt; RR2 &lt; range RR2)</td>
</tr>
<tr>
<td>Rule 4</td>
<td>(range RR1 &lt; RR2 &lt; range RR1) &amp; (range RR1 &lt; RR2 &lt; range RR2)</td>
</tr>
</tbody>
</table>

TABLE 2. N, S, V additional rule

<table>
<thead>
<tr>
<th>Rule</th>
<th>Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>N Rule 5</td>
<td>(0.8 &lt; RR2) &amp; (0.8 &lt; RR2) &amp; (0.9 &lt; RR2) &amp; (0.9 &lt; RR2)</td>
</tr>
<tr>
<td>S Rule 5</td>
<td>(median RR1 &lt; RR2) &amp; (RR1 &lt; 1.0)</td>
</tr>
<tr>
<td>V Rule 5</td>
<td>(0.2 &lt; RR2 - RR1) &amp; (0.2 &lt; RR2 - RR1)</td>
</tr>
</tbody>
</table>

III. EXPERIMENTAL RESULTS

Fig. 2 shows the results of pre-processing that applies normalization and DWT to the original signal. Fig. 2(a) shows the original signal, whereas Fig. 2(b) shows the results of the application of normalization and DWT that adjust the ratio of amplitude and the baseline and remove unnecessary noises. A summary of all results is shown in Table 3.

IV. CONCLUSION

In the present study, an arrhythmia classification method was proposed that combined features extracted through PCA and LDA and used fitness rules and the KNN algorithm based on the combined features. The proposed algorithm produces slightly better results. Future studies to develop pre-processing and classification optimization algorithms will be necessary to further increase the classification speeds and enhance classification accuracy.

REFERENCES

Speeding Up SVM Training in Brain-Computer Interfaces

David Lee
Department of Digital Media, Media Engineering
The Catholic University of Korea
Bucheon, Gyeonggi, South Korea

Woo-Hyuk Jung, Sang-Goog Lee
Department of Media Technology & Media Contents
The Catholic University of Korea
Bucheon, Gyeonggi, South Korea

Abstract—Traditional Support Vector Machine (SVM) is widely used classification method for brain-computer interface (BCI). However, SVM has a high computational complexity. In this paper, Gaussian Mixture Model (GMM)-based training data reduction is proposed to reduce high computational complexity. The proposed method is configured as follows: First, wavelet-based combined feature vectors are applied for motor imagery electroencephalography (EEG) identification and principal component analysis (PCA) are used to reduce the dimension of feature vectors. Thereafter, the GMM is implemented to reduce training data sizes. Finally, a nonlinear SVM classifier is used to classify the reduced training data. The performance of the proposed method was evaluated using three different motor imagery datasets in terms of accuracy. The results from the study indicate that the proposed method achieves high accuracy with faster computational time in motor imagery EEG classification.

Keywords—brain-computer interface; wavelet transform; training data reduction; support vector machine

I. INTRODUCTION

BCI refers to the technology that analyzes humans’ mental activities to enable the brain to issue orders directly to computers [1]. Therefore, BCI can help not only general users but also physically challenged persons that cannot move their muscles due to neurological abnormalities and the elderly and infirm that are restricted in movements in many areas. To this end, BCI should improve the trade-off between accuracy and speed or strike a balance between accuracy and speed [2].

To improve BCI’s classification accuracy, recent EEG based BCI studies reported investigations and evaluations of diverse classification algorithms [3-5]. Among them, many researchers studied motor imagery EEG signal classification using SVMs that show a good generalization of binary classification by maximizing the margin existing between two data classes. However, SVMs have problems with computational time. The training process should solve quadratic optimization problems and the testing process relies on the number of support vectors generated in the training process.

This study investigates a method to improve SVM training speeds of motor imagery EEG by reducing combined feature vector based training data using the GMM method. We evaluate the proposed method using the accuracy, kappa, mutual information (MI), and computation time to see how efficient the performance of the proposed method is.

II. MATERIALS AND METHODS

A. Datasets

The three datasets, recorded by the Graz BCI group, used to evaluate the performance of the proposed method. All these three datasets were obtained while the test subjects were performing left hand and right hand motor imagery.

B. Feature Extraction

During motor imagery work, special features termed Event Related Desynchronization (ERD) and Synchronization (ERS) are generated from mu waves and beta waves throughout the entire area of the sensory motor cortex. In many BCI studies, such ERD/ERS components appearing from mu and beta waves were used as features of EEG signals for identification of motor imagery. In the present study, not only the 8-30 Hz band that includes mu and beta waves but also some theta wave bands are also used because theta waves also include EEG features for identification of motor imagery EEG signals. To analyze EEG signals, continuous wavelet transform (CWT) and discrete wavelet transform (DWT) have been used for feature extraction.

C. Combined Feature Vector

During In the present study, the features extracted as such are combined into one single feature vector. This method can obtain macroscopic and microscopic information on time and frequency to improve the accuracy of motor imagery classification. The combined feature vector consists of a total of 96 features:

\[ \text{Combined feature} = \{C_1, \ldots, C_{30}, D_1, \ldots, D_{36}\} \] (1)

D. Feature Selection

Feature selection is a process to select optimum features in high-dimensional feature spaces. Since we may be put under a dimensionality curse if the number of training data is smaller than the number of features, a process to select features that have important information is essential. Among feature selection methods, principal component analysis (PCA) has been widely used in diverse areas. The primary purpose of
Beyond PACS2016 existing training data, we selected the mean vectors $\mu$ was improved by 0.3 -8.1% than the other methods for most results showed that the performance of the proposed method achieved better classification accuracy by approximately 2.4% on average when the entire training data were used and by approximately 0.8% when 50% of the entire training data were used.

E. Training data Size Reduction

The GMM is a generative model for clustering and probability densities estimation using statistical methods. Since it also uses covariance matrices, it includes information on the distribution of features. Thus, the GMM conducts modeling using M component Gaussian densities for the distribution of training data $x$. In the present study, to reduce the sizes of existing training data, we selected the mean vectors $\mu_i$ of M Component Gaussian densities as representative training data of individual classes. Because M represents the size of training data for each class, the size of training data for all classes becomes $2\times M$. In the experiment, $2\times M$ was set as 5, 10,..., 100% of the entire training data.

F. Classification

The SVM is a discriminative model that has been widely used for linear and nonlinear classifications due to its high prediction accuracy. The SVM determines the optimal hyperplane (i.e. decision boundary) that has the largest margin between two classes. Data from classes closest to a decision boundary are called support vectors. In the present study, radial basis kernel functions were used as SVM’s kernel functions:

$$K(x_i, x) = \exp \left( -\frac{||x_i - x||^2}{2\sigma^2} \right)$$  \hspace{1cm} (2)

where $\sigma$ is a kernel parameter that is related kernel widths. In the present study, cross-validation (CV) was applied to evaluate and select the optimal values of C and $\sigma$ through the grid search.6 The range of parameter was given as follows: C=2$^{-5}$, ... , 2$^{15}$, $\sigma=2^{-15}$, ... , 2$^{-1}$. The fold number CV was set to 10.

III. EXPERIMENTAL RESULTS

First, we intend to demonstrate the efficiency of combined features vector through comparison between DWT, CWT, and the combined feature vector for the three datasets. Thereafter, to demonstrate the efficiency of the GMM used to reduce size of training data, we compare the performance of the GMM with that of the random sampling method. Since training data reduced through the GMM are generated differently every time, these processes were repeated 10 times in order to obtain reliable evaluation results.

Finally, complete content and organizational editing before formatting. Please take note of the following items when proofreading spelling and grammar:

A. Performance of Training data Size Reduction

The proposed method showed better performance than the other methods. When the entire training data were used, the results showed that the performance of the proposed method was improved by 0.3-8.1% than the other methods for most subjects except for some subjects (subjects 4 and 7). When only 50% of the entire training data were used, the results showed that the proposed method achieved better classification accuracy than the other methods in all subjects except 4, 5, 7, and 9. Consequently, the combined feature vectors improved classification accuracy by approximately 2.4% on average when the entire training data were used and by approximately 0.8% when 50% of the entire training data were used.

### Table I. Comparative Results of the Feature Extraction Methods (DWT, CWT, Whole Training Data and Combined Feature Vectors Using About 50% of the Whole Training Data) in Terms of the Average Classification Accuracy (%). Bold when Classification Accuracy is Highest.

<table>
<thead>
<tr>
<th>Subjects</th>
<th>DWT</th>
<th>CWT</th>
<th>ALL-Combined Feature</th>
<th>50%-Combined Feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>92.9</td>
<td>94.1</td>
<td>98.5</td>
<td>95.1</td>
</tr>
<tr>
<td>2</td>
<td>73.3</td>
<td>81.4</td>
<td>83.1</td>
<td>81.9</td>
</tr>
<tr>
<td>3</td>
<td>92.5</td>
<td>80.7</td>
<td>83.4</td>
<td>82.3</td>
</tr>
<tr>
<td>4</td>
<td>75.2</td>
<td>77.7</td>
<td>76.7</td>
<td>75.3</td>
</tr>
<tr>
<td>5</td>
<td>52.6</td>
<td>60.1</td>
<td>61.8</td>
<td>58.9</td>
</tr>
<tr>
<td>6</td>
<td>55.4</td>
<td>55.6</td>
<td>56.2</td>
<td>55.9</td>
</tr>
<tr>
<td>7</td>
<td>95.5</td>
<td>96.2</td>
<td>96.1</td>
<td>95.2</td>
</tr>
<tr>
<td>8</td>
<td>88.3</td>
<td>87.4</td>
<td>94.1</td>
<td>92.7</td>
</tr>
<tr>
<td>9</td>
<td>79.2</td>
<td>87.8</td>
<td>88.1</td>
<td>87.1</td>
</tr>
<tr>
<td>10</td>
<td>74.8</td>
<td>72.6</td>
<td>80.7</td>
<td>77.6</td>
</tr>
<tr>
<td>11</td>
<td>90.6</td>
<td>87.7</td>
<td>90.0</td>
<td>89.1</td>
</tr>
<tr>
<td>12</td>
<td>78.7</td>
<td>83.3</td>
<td>84.8</td>
<td>83.3</td>
</tr>
<tr>
<td>Avg.</td>
<td>76.3</td>
<td>80.4</td>
<td>82.8</td>
<td>81.2</td>
</tr>
</tbody>
</table>

IV. CONCLUSION

The present paper proposed DWT and CWT based combined feature vectors to improve SVM based classification accuracy in the field of BCI applications. In addition, the present paper proposed a GMM based training data reduction method to improve the speed of SVM training processes. The performance of the proposed method was evaluated in terms of accuracy, MI, kappa, and Computation time using diverse BCI datasets. The proposed method considerably reduced the size of entire training data while achieving small losses (1%). Consequently, the number of selected SVs decreased and SVM training speed increased by up to eight times. In addition, since the number of SV decreased, testing speed was improved.

### REFERENCES

An Analysis on Variability in Motor Learning of Patients with Parkinson’s Disease through a Virtual Throwing Task

Jinseok Oh1 Je-Kwang Ryu2* Se-Woong Park3 Kyung-Min Lee1,4
1. Program in Cognitive Science, Seoul National University, Seoul, Korea
2. Institute for Cognitive Science, Seoul National University, Seoul, Korea
3. Department of Biology, Northeastern University, Boston, MA, USA
4. Department of Neurology, Seoul National University Hospital, Seoul, Korea
* kidd@snu.ac.kr

Abstract—Parkinson’s disease is well-known for its degenerative effect on the human motor system. Patients with the disease show signs of deficiency in motor performance. Moreover, reduced level of dopamine release, a fundamental characteristic of the disease, seems to influence not only the learning process of a new motor skill but also the retention of the learned content. Hence, it has been controversial whether patients with Parkinson’s disease can successfully acquire a new motor skill and maintain it for a long time. Based on a new analysis technique that investigates variability in motor learning, we argue that the patients are indeed capable of obtaining a new motor skill to a degree through sets of practices and maintain it.

Keywords—motor learning, variability, noise, Parkinson’s disease, implicit learning, skill acquisition, TNC-Cost Analysis

I. INTRODUCTION

Parkinson’s disease is a neurodegenerative disorder characterized by bradykinesia, resting tremor, postural instability, and autonomic features [1]. The disease is believed to be caused by the degeneration of the dopaminergic nigrostriatal system [1], or cells in substantia nigra pars compacta in particular. Reduced amount of secreted dopamine, disrupts the normal activity of basal ganglia. When the basal ganglia behave atypically, both the selection of proper behaviors and the inhibition of unwanted ones are executed poorly.

Motor learning, and motor skill learning, in particular is classically defined as a set of processes that takes place during practice or experience and leads to relatively permanent changes in the capability for movement. Prominent changes include the increased accuracy of movements with practice. Specifically, motor skill learning can be done in two manners: explicitly and implicitly. The difference lies in the consciously accessible knowledge of what has been learned. An example of implicit learning could be learning how to ride a bicycle.

Nigrostriatal system and related brain areas such as the striatum and the basal ganglia are closely associated with implicit skill learning [2,3]. People with Parkinson’s disease, who have damaged basal ganglia or striatum, show the sign of reduced, or no learning of new motor skills [4]. Moreover, brain imaging studies have revealed that basal ganglia and other related brain regions are more activated when learning a new motor sequence [5]. However, other researchers have argued that people with Parkinson’s disease are still able to learn new motor skills and acquire it for a long time after learning finishes [6,7]. The discordance of the studies arises mainly for two reasons. Above all, the tasks the researchers used were too simple, and not much motor learning oriented. Rather, tasks required participants to implicitly adapt to seemingly random perceptual sequences. Motor skills such as finger tapping were mere tools to represent the participants’ level of adaptation to given sequences. Furthermore, conventional methods of determining whether learning was successful have concentrated on comparing the learning outcome with the initial condition, and not on addressing the specific changes within the learning process. In order to resolve this issue, both a task designed to measure motor learning per se and a method of quantifying the relevant changes during learning should be prepared.

Variability in motor learning could be one relevant change that can be quantified. Among several causes of variability, an inherent and ubiquitous neuro-motor noise is a prominent one. Even performing the same movement twice does not guarantee the identity of the two different movements, as the two are contaminated separately by the noise in planning as well as in executing movements. Redundancy may be another significant cause of variability. Motor tasks are often redundant in that numerous different ways to achieve goals exist. In such cases, even the two exactly same outcomes can be reached by distinct combinations of movements. Therefore, while it is known that the amount of variability in performance during motor learning decreases when learning continues and skills improve [8], careful examination on which kind of variability is reduced at which moment is required to fully understand what is going on during motor skill learning. A recent attempt named TNC-cost analysis method [8] separates variability in motor learning into three categories of costs - tolerance, noise, and covariation – and quantifies each cost to observe how a whole learning process could be analyzed in relation to the changes of the amounts of the costs. With the aid of this analysis technique as well as a task devised to study motor learning, the present study aims to see the variability in motor learning of patients with Parkinson’s disease. The study intends to address two questions. Are patients with Parkinson’s disease still capable of attaining new motor skills? If so, would the learning process of the patients be any different from that of normal individuals? Discussing the results of the study is expected to provide some helpful insight about how nigrostriatal system and related brain areas contribute to motor skill learning.
II. METHOD

A. Participants

A total of 31 participants were recruited. 22 were patients with mild to moderate Parkinson’s disease (11 females, mean age = 61.4, sd = 5.54) and 9 were normal individuals with no cognitive and motor impairment (5 females, mean age = 61.6, sd = 5.64).

B. Task

A simplified and virtual version of the British pub game, Skittles [9], was used. Participants were instructed to throw a ball using an experiment device while sitting to hit a target on a TV screen. Participants practiced 4 sessions over 10 days (3 practice sessions in a row, and the retention session a week after the third session). Each session had 6 blocks of 60 trials.

The task was able to limit the participant’s movement to a horizontal rotation done in a two-dimensional space. Thus, the performance was determined by two independent, or execution variables: angle and angular velocity at release. The dependent or outcome variable was the minimum distance between the center of the target and the released ball’s trajectory (error). The error was set to 1 whenever the ball hit the center post. The task was redundant in that hitting the target could be done by numerous distinct combinations of the execution variables.

C. Analysis

The TNC-cost analysis was applied to see how different types of variability in motor performance changed over practice sessions. In addition, trajectories, as well as error plots, were drawn to directly observe the gradual progress of the participants’ performance.

III. RESULT

A. Release Angle and Release Velocity

Both the patients and the control participants were similar in terms of how the mean release angle changed over time. Exploring from 60 to 120 degree, most participants terminated the retention sessions at around 90 degrees from the starting point. Groups were, however, dissimilar in the change of the mean release velocity over time. While the control participants started around 250 deg/s and maintained consistent mean release angle and velocity were chosen more frequently. It can be argued that patients with Parkinson’s disease can indeed learn a new motor skill, based on the fact that their performances at the retention session were similar to those at the last practice session, and that they could reduce neuro-motor noise, or variability in performance, to the same level as the normal people did. That the patients are slower in finding appropriate combinations of release angles and velocities can be attributed to their lack of motor control. Nevertheless, a possibility that such deficit in motor control might affect the patients’ cognitive activity of solution finding and further delay learning a new motor skill remains. Future studies are expected to reveal the putative role the level of motor control plays in shaping the cognitive change in the process of motor learning.

REFERENCES

Visual Imagination from Texts

Hanock Kwak and Byoung-Tak Zhang
School of Computer Science and Engineering
Seoul National University
Seoul 151-744, Korea
Email: (hnkwak, btzhang)@bi.snu.ac.kr

Abstract—Imagination is a fundamental ability of humans which resides in the cognitive system. We propose a connectionist model that generates images from a given sentence after trained on a dataset of image-sentence pairs. The model is composed of language model and image model that are connected with a latent variable constrained by a prior distribution. The latent variable encodes dual information and it is generalized by Bayesian learning method. We trained on cartoon video series ‘Pororo’ and 16,066 fine-grained sentences describing short clips. Our model successfully generates plausible images which are highly correlated with a given sentence.

I. INTRODUCTION

Images are composed of several different objects forming a hierarchical structure with various styles and shapes. Deep learning models are used to disentangle those complex underlying patterns [1][2], build distributed feature representations [3], and solve classification [4] and generation [5] problems using large datasets. The objective of model governs way of encoding complex visual informations as well as decoding latent variables.

Natural language offers a general, abstract and flexible descriptions of visual informations. Even language lacks detailed visual properties humans are able to imagine specific objects from abstract information depending on the experiences of what they seen before. Traditionally visual information about an object has been captured in attribute representations[6][7]. The categories of images are most simple one, and descriptions are more general. We connect this abstract representations with detailed visual informations generalized in real vectors.

Recently, deep convolutional and recurrent networks for text and image have successfully learned discriminative and generalizable representations automatically from raw data[8]. These approaches exceed the previous state-of-the-art methods that do not use neural networks. Inspired by these works, our model learns a direct mapping from sentences to image pixels using recent deep learning techniques.

II. MODEL

The model is combination of language model and image model that are connected with a latent variable constrained by a prior distribution. The language model is mainly made of LSTM [9] which encodes a given sentence and the image model uses transposed convolutional layers[5] that generate images from encoded vector. The model is illustrated in Figure 1.

Fig. 1. Illustration of the model. Firstly, a sentence is passed in LSTM recurrently word by word. Then the final output of the LSTM is given to transposed convolutional layers which generates an image.

Let a given sentence be a vector \( s = [w_1, w_2, ..., w_n]^T \) where \( w_i \) is a \( i \)-th word. Then a representation \( z \) of the sentence \( s \) is a final output vector of LSTM:

\[
h_t = LSTM(h_{t-1}, Uw_i), \quad z = h_n. \tag{1}
\]

where \( h_t \) is a hidden state of LSTM at \( t \) step, and \( U \) is an embedding matrix. The final output image \( y \) is generated from \( z \) through transposed convolutional layers. We used mean square error between \( y \) and ground truth image as a loss function.

In addition, several recent deep learning techniques (batch normalization [10], ADAM [11], etc.), critical to the overall performance, were utilized.

III. EXPERIMENTS

We trained on cartoon video series ‘Pororo’ and 16,066 fine-grained sentences describing short clips. The ‘Pororo’ has long play time of 1,232 minutes, various scenes, and some characters. All images are resized to 64 x 64 with antialiasing. The image model has series of four transposed convolution where channels are halved and size of the filter maps doubled for each time. Initially we set hidden state \( h_1 \) as zero vector and \( U \) is initialized with pre-trained Glove vectors[12].

The result is shown in Figure 2. The model successfully generated plausible images from sentences and it generalized dual representation of texts and images.
Fig. 2. Some samples of output images.

IV. Conclusion

The model generalized dual representation and connected both modality to generate images from sentences. Yet output image is not clear enough and the connection is ambiguous. We need to design models that can perform more accurate and logical inference between texts and images.

REFERENCES


KAP Based PICO-Compliant Query Construction

Muhammad Afzal
Department of Computer Science & Engineering,
Kyung Hee University (KHU),
Yongin, South Korea
muhammad.afzal@oslab.khu.ac.kr

Sungyoung Lee
Department of Computer Science & Engineering,
Kyung Hee University (KHU),
Yongin, South Korea
sylee@oslab.khu.ac.kr

Abstract—In recent years, the volume of biomedical research literature has grown exponentially which posed challenges for the clinical practitioners in terms of “well-built” question preparation and search query generation. The key objective is to facilitate the process of automatically constructing a “well-built” question in order to save preventable time.

Keywords—PICO; Query; Mapping Model; KAP; MAP.

I. INTRODUCTION

In healthcare domain while practicing evidence-base medicine, formulating a well-built question becomes a necessity. It poses several challenges such as how to formulate the question? From where to populate the contents for the question? Formulating a question in PICO format simplifies the next steps of searching from online literature. Theoretically, it is an appealing form to make an appropriate question for searching, however, the big hurdle is to provide the adequate contents for different parts of PICO. Doing manually, it become time consuming for the busy clinical practitioners. At the same time, it is hard for a non-expert practitioners to comprehend the real meaning of each part without having a deep understanding of the domain concepts with respect to PICO. In order to make the question construction easier for a user, we need an automatic mechanism for content filling in PICO template.

II. METHODS

We propose a generic model called (Knowledge Alignment to PICO (KAP) that provides two level mappings: structure- and concepts-level mappings.

A. Structure-level mappings

The proposed KAP model provides structure mapping guidelines of a clinical decision support system (CDSS) knowledge to map with PICO. KAP model is independent of any specific knowledge representation formalism thus provides the flexibility to be used for different knowledge representations. In order to realize the KAP model, we propose specialized mapping models such as MAP (MLM Alignment to PICO), GLAP (GLIF Alignment to PICO), PRAP (Production Rule Alignment to PICO), and others. The specialized models map the slots of a particular knowledge representation scheme physically.

B. Concepts-level mappings

At concept level mappings, the concepts are matched with available standard terminologies such as SNOMED CT and UMLS. Once a concept is matched, it is further confirmed whether the matched concept belongs to any specified part of PICO or not. For a correct matched concept, it is determined whether it has numeric value or descriptive. Concepts having numeric values are given comparatively less importance due to their impact on the hit ratio in information retrieval. Finally, the salient concepts are concatenated to construct the final PICO-compliant query.

III. CONCLUSION

In this work, we introduced the concept of automatically constructing a PICO-compliant query from diverse sources of knowledge representation. We used two level mappings: structure- and concepts-level. This work opens future research venues such as validation and verification for a fully automated query in order to make it more target oriented.

ACKNOWLEDGMENT

This work was supported by the Industrial Core Technology Development Program (10049079, Develop of mining core technology exploiting personal big data) funded by the Ministry of Trade, Industry and Energy (MOTIE, Korea).
Model-Driven Adaptive User Interface based on Context and User Experience Evaluation

Jamil Hussain
Department of Computer Engineering Kyung Hee University Seocheon-dong, Giheung-gu, Yongin-si, Gyeonggi-do, Republic of Korea.
jamil@oslab.khu.ac.kr

Sungyoung Lee
Department of Computer Engineering Kyung Hee University Seocheon-dong, Giheung-gu, Yongin-si, Gyeonggi-do, Republic of Korea.
sylee@oslab.khu.ac.kr

Abstract— Personalized services have greater impact on user experience to effect the level of user satisfaction. Many approaches provide personalized services in the form of an adaptive user interface. The focus of these approaches is limited to specific domains rather than a generalized approach applicable to every domain. In this paper, we proposed a domain and device-independent model-driven adaptive user interfacing methodology. Unlike state-of-the-art approaches, the proposed methodology dependent on the evaluation of user context and user experience (UX). The proposed methodology is capable of adapting user interface based on the utilization of contextual factors at runtime using the adaptation rules.

Keywords— Human Computer Interaction, Personalized user interface, Adaptive user interface, User Experience, Context-aware user interfaces

I. INTRODUCTION

With the advancement of technology, Human Computer Interaction (HCI) plays very important role. Among other, user interface (UI) is a dominant part of interactive systems. End user applications provide many functionalities, which increases the complexities of UIs. Most of the users use a small portion of the offered functionality and major part goes underutilized. Additionally, the UI element usage are differ among users.

The personalization in the user interface can be introduced in two ways: adaptable and adaptive. Both ways can improve the positive user experience in terms of usability and user satisfaction as compared to a default user interface. In the adaptable approach, the personalization made by the user, while in the adaptive approach, the personalization is achieved by system. Nevertheless, the adaptable approach requires frequent customizations according to the user’s skills. Many of the users are not able to personalize the UIs according to their preferences due to their skills. Additionally, the adaptive personalization extends the system lifespan, enable the users to achieve their goal, and increases the operational accuracy and speed.

In this paper, we propose an adaptive UI methodology that tailors the UI on the basis of evaluation of user context and user experience. The main objective of the proposed methodology is to deal with the personalized approach towards building and managing the user interfaces.

II. ADAPTIVE UI GENERATION METHODOLOGY

The real adaptive behavior in an adaptive user interface process is shown in Figure. 1, which starts from user interaction with the system. The monitoring module is responsible for data collection while user is interacting with the system through different sensors and trackers (e.g., facial, vocal, eye, and analytics). We also consider the user feedback as self-reported data. The evaluator component evaluates the acquired information and decides whether adaption is required on UI using adaptation rules or not. It also evaluates the user experience on periodic bases depend on the configuration. The user experience is measured using UX metrics such as performance, self-reported, and physiological metrics based on the collected data (explicit and implicit) through the monitoring modules. If any adaptation is needed, UI is adapted accordingly, otherwise ignored.

Figure 1. Adaptive Behavior Data Flow

III. CONCLUSION

In this paper, we present the Adaptive UI methodology that adapts the user interface using model-driven approach based on user preferences and their experience. The proposed approach considers the dynamics of the UI associated with the user in the form of context-of-use. It helps in improving the information accessibility, usability, and user experience of system.

ACKNOWLEDGMENT

This work was supported by the Industrial Core Technology Development Program (10049079, Develop of mining core technology exploiting personal big data) funded by the Ministry of Trade, Industry and Energy (MOTIE, Korea).
IoT based Human Centric Context Awareness Framework for Healthcare and Wellness Platform


*Department of Computer Engineering, Kyung Hee University, South Korea
†Telemedicine Group, Center for Telematics and Information Technology, University of Twente, Netherlands
Email: ∗{ jhb, dwkang, hth, thienht, asif.razzaq, wajahat.alkhan, sylee}@oslab.khu.ac.kr, † o.banoslegran@utwente.nl

Abstract— In recent years, healthcare and wellness platforms are developed rapidly with the advent of smart devices which possess diverse sensors. Existing systems are limited to provide simple health status visualization services from single device or single sensor, which make them unable to provide timely high quality services. This paper proposes Human Centric Awareness Framework based on diverse IoT devices to provide high quality services timely.

Keywords— human behavior; context awareness; machine learning; ontology

I. BACKGROUND
Context aware technologies are advancing and therefore leading to the increase in services built on top of it. Mining Minds platform [1] is one such platform that is utilizing the recognized context of the user to provide diverse services. Context recognition process is handled at the Information Curation Layer (ICL) [2], that divides it into low level and high level context. Low level context includes the context recognized from the sensory devices, smartwatch, camera, and smartphone. The low level context is based on activities (walking, running, sitting, and others), location (home, office, gym, and others), emotion (happy, sad, anger, and others), and food (rice, burger, meat, and others) recognition. These high level activities are utilized by the high level context (amusement, exercise, commuting, and others in activities [3], and grain, carbohydrate, fats and others in nutrition services) to recognize context on top of the low level context. Both the high level and low level context is used by the Service Curation Layer (SCL) for the generation of the services.

II. HUMAN CENTRIC CONTEXT AWARENESS FRAMEWORK
The proposed Human Centric Context Awareness Framework is hierarchically structured model able to recognize low-level (LLCA) and high -level context (HLCA) with heterogeneous sensor data such as smartphone, smartwatch and 2D camera. LLCA recognizes various human real-life context such as physical activities, emotion status, locations and food intake from various sensors. HLCA recognizes accurate high-level context based ontological model with extracted low-level context. LLCA recognize 9 physical activities, 4 emotions and 5 locations and HLCA recognize 9 behavioral context and 3 nutrition context.

III. CONCLUSION AND FUTURE WORK
This paper proposes IoT based Human Centric Context Awareness Framework by using various devices and sources in real-time manner. In the future, we are planning to enhance accuracy of existing modules and extend the number of input d, such as physiological sensor and smart cup.

ACKNOWLEDGMENT
This work was supported by the Industrial Core Technology Development Program (10049079, Develop of mining core technology exploiting personal big data) funded by the Ministry of Trade, Industry and Energy (MOTIE, Korea).

REFERENCES
Service Curation Framework for Context-Aware Personalized Recommendations

Syed Imran Ali, Muhammad Sadiq  
Department of Computer Science and Engineering 
Kyung Hee University, Seocheon-dong, Giheung-gu, Yongin-si, Gyeonggi-do, Republic of Korea, 446-701 
{Imran.ali,sadiq}@oslab.khu.ac.kr

Muhammad Afzal, Sungyoung Lee  
Department of Computer Science and Engineering 
Kyung Hee University, Seocheon-dong, Giheung-gu, Yongin-si, Gyeonggi-do, Republic of Korea, 446-701  
{Muhammad.afzal, sylee}@oslab.khu.ac.kr

Abstract—Self-quantification has gained a lot of traction in wellness-savvy individuals. A huge number of systems are introduced in order to meet this demand. The core features of most of these systems consist of self-tracking, self-quantification, and raw sensory data based self-monitoring. Such powerful features provide users an enhanced awareness about their quotidian activities in a consistent and timely manner. As an extension, we propose a comprehensive system which is capable to actively induce healthy behavior in users by providing accurate, context-aware and timely recommendations by processing user’s contextual information. A novel methodology is proposed for acquiring, processing and interpreting context-aware personalized recommendations.

Keywords—context-aware; personalized; recommendation;

I. INTRODUCTION

The current advancements in systems for personalized healthcare and wellness monitoring have created a new market for wellness savvy individuals. These systems can monitor a person’s daily activities, eating habits and can provide goal-based recommendations. Although the data which is generated by a user is efficiently processed for generating a recommendation however these systems fall short of inducing healthy habits in users due to their inability to account for contextual information. Previously, we proposed Mining Minds framework for personalized healthcare and wellness support. It utilizes expert knowledge and contextual nuggets of information for providing personalized recommendations. The proposed system provides both physical activity and nutrition-based food recommendations.

II. PROPOSED METHODOLOGY

The high level architecture of the proposed system is depicted in Fig. 1.

Service Orchestrator (SO) serves as the gateway to SCF. Data is collected from wearable sensors including smart phone and subsequently curated and stored with a logical division of user profile data, contextual data, and environmental data.

Recommendation Builder (RB) component performs inferences over rules stores in the knowledge base, user profile data, contextual data, and environmental data.

Recommendation Interpreter (RI) in Fig. 1, gets the initial recommendation built by recommendation builder component and interprets it contextually. In order to achieve this task a comprehensive survey is conducted in which physical activity based preferences are collected from a diverse population set. For example RI decides whether a user can be interrupted for performing a physical activity (e.g. running) based on the user’s contextual information (e.g. location/office) or not. Moreover it is also analyzed whether the generated recommendation (by RB) is contextually suitable for the user or not (e.g. recommended running outside while it is raining). And if possible to provide with a more amenable alternative recommendation.

Context-aware personalized recommendation is the key contribution of the proposed system.

III. CONCLUSION

Service curation framework is designed to coordinate services for providing personalized recommendations. This work is carried out as a part of Mining Minds project which is a comprehensive digital framework for health and wellness services.

ACKNOWLEDGEMENT

This work was supported by the Industrial Core Technology Development Program (10049079), Develop of mining core technology exploiting personal big data) funded by the Ministry of Trade, Industry and Energy (MOTIE, Korea).
Smart CDSS: Knowledge Acquisition and Modeling using Mind Maps and Decision Trees

Musarrat Hussain, Maqbool Hussain
Department of computer science and engineering, Kyung Hee University, Yongin, South Korea, {musarrat.hussain, maqbool.hussain}@oslab.khu.ac.kr

Sungyoung Lee
Department of computer science and engineering, Kyung Hee University, Yongin, South Korea, sylee@oslab.khu.ac.kr

Abstract—Clinical practice guidelines (CPGs) is one of the key knowledge resources used in medical domain. CPGs are mainly available in an un-structural and a semi-structural form. For a concrete knowledge, domain expert rigorously investigates the CPGs and convert them into a human readable and computer interpretable format. In this paper, we demonstrate knowledge acquisition and its modeling from the oral cavity cancer guidelines using Mind Maps and Decision Trees (DTs).

Keywords—Mind Maps; Decision Trees; Clinical Practice Guidelines; Clinical Decision Support System

I. INTRODUCTION

A CDSS plays a pivotal role in healthcare while improving the patient care and reducing the chance of errors. The key quality of CDSS services is based on a well-represented knowledge supported by trusted knowledge resources. CPGs are considered trusted in medical domain. However, CPGs in published form are not computer interpretable due to its unstructured format. This paper demonstrates the method of representing CPGs in Mind Maps and conversion to DTs.

II. METHOD

A. Investigating CPGs and developing Mind Maps

The first most important step in knowledge creation is to define the objectives of CDSS intervention and select appropriate CPGs. Second, while targeting the intended objectives, the structural CPGs are investigated by the domain experts. The initial sketch is drawn with high level concepts represented as “Central Topics”, “Main Topics”, and “Topics” with appropriate relationships. Third, the domain expert finalizes the model into a refined Mind Map.

B. Conversion of Mind Maps to Decision Trees

Mind maps are human understandable but represent the knowledge only at a high level. To make the knowledge explicit, a conversion from Mind Maps to a concrete knowledge representation such as DTs is required. The conversion process includes: identifying and isolating candidate conditions and actions, and drawing appropriate branches to reflect the actual knowledge.

III. RESULTS: CASE STUDY

This work is carried out as a part of Smart CDSS in the area of head and neck cancer. Figure 1 & 2 depict the partial knowledge representation for oral cavity cancer.

IV. CONCLUSION

This paper proposes the knowledge acquisition for CDSS from CPGs using Mind Maps and DTs. The method is demonstrate with oral cavity cancer and the knowledge is used in a real project of Smart CDSS.

ACKNOWLEDGMENT

This work was supported by the Industrial Core Technology Development Program (10049079, Develop of mining core technology exploiting personal big data) funded by the Ministry of Trade, Industry and Energy (MOTIE, Korea).
Semantic Reconciliation Model for interoperable and shareable knowledge authoring environment

Taqdir Ali
Ubiquitous Computing Lab, Department of computer science and engineering, Kyung Hee University, South Korea, taqdir.ali@oslab.khu.ac.kr

Sungyoung Lee
Ubiquitous Computing Lab, Department of computer science and engineering, Kyung Hee University, South Korea, sylee@oslab.khu.ac.kr

Abstract—The intelligent recommendations, generated by a Clinical Decision Support System (CDSS), depend on up-to-date knowledge base. An interoperable and a shareable knowledge base reduces the burden of domain experts to transform their heuristics and experiences. The goals of interoperability and shareability of knowledge can be achieved using standard data models and standard terminologies. We proposed a mapping model called “semantic reconciliation model” to provide mappings among standard terminologies, standard data model, and domain clinical model.

Keywords—Clinical decision support system; interoperable knowledge; shareable knowledge; semantic reconciliation

I. INTRODUCTION
Healthcare standard terminologies, a standard data model, and a standard knowledge representation play a vital role in creating shareable and interoperable knowledge base. HL7 Virtual Medical Record (vMR) provides a standard information model to represent clinical information for scalable and interoperable clinical decision support systems. Likewise, standard terminologies enhance the interoperability of the knowledge base. The shareable knowledge creation in the form of standard representation of HL7 Medical Logic Module (MLM) increases the knowledge acquisition complexity with respect to usability. Therefore, our proposed methodology, is called semantic reconciliation model (SRM), provides multi-model mappings among standard terminologies, a standard data model, and a domain clinical model (DCM) to hide the complexity of standard knowledge creation. The system transforms the plain rules into a shareable MLM representation with the combination of standard terminology (SNOMED CT) and standard data model (vMR) in an automatic manner. Therefore, the standard MLM representation of knowledge achieves the shareability while vMR, SNOMED CT, and DCM combined with MLM enhance the interoperability of knowledge.

II. METHODOLOGY
The proposed system takes DCM, SNOMED CT, and vMR ontologies and specification to generate three mapping files DCM-SNOMED, vMR-SNOMED, and DCM-vMR. In DCM and SNOMED CT mappings we use different mapping algorithms like string matching, synonym matching, labels matching, child matching, and property matching. The proposed system provides choice for a subset of algorithms execution or complete set of algorithms execution. In second mapping, i.e. vMR-SNOMED CT, we use different similarity matching algorithms like Jaccard, Euclidean, and cosine algorithms. At the end of execution of similarity algorithm, the system decides the final matching concept based on maximum similarity technique. Once, we get DCM-SNOMD mapping and vMR-SNOMED mapping then the third mappings of DCM-vMR achieved by transitivity law. The three resultant mapping files are used as an input to the Automatic Shareable Knowledge Creation. This module generates the MLM from the plain rule through the mappings generated by SRM. The MLM is generated on the basis of its standard structures of different slots of “maintenance”, “library”, and “knowledge”. The SRM mappings mainly play role in transformation of “knowledge” slot. The overall methodology of SRM is depicted in Fig.1

III. CONCLUSION
The semantic reconciliation model (SRM) is developed for interoperability and easy integration of shareable knowledge base with clinical workflows. The complexity of shareable knowledge creation is hidden with automatic generation of MLM.

Acknowledgement: This work was supported by the Industrial Core Technology Development Program (10049079, Develop of mining core technology exploiting personal big data) funded by the Ministry of Trade, Industry and Energy (MOTIE, Korea).
UA V-assisted Intelligent Crowdsourcing in Natural Calamity

Sarder Fakrul Abedin, Md Golam Rabiul Alam, Anupam Kumar Bairagi, Ashis Talukder and Choong Seon Hong
Department of Computer Engineering, Kyung Hee University, South Korea
Email: {saab0015, robi, anupam, ashis, cshong}@khu.ac.kr

Abstract—During the natural calamity, it becomes challenging to coordinate the search and rescue operation due to the possible destruction of the static network infrastructure and the lack of sensory observation data support to localize the affected area and victims. Therefore, in this research, we have focused on solving the disaster time intelligent crowdsourcing through UAVs in an energy efficient way for victim localization provided with a quick response to conduct the emergency rescue operation. We have model the energy efficient crowdsourcing through $\epsilon$-greedy Q-learning and the simulation result shows the efficiency and convergence of the modeled algorithm for the proposed energy efficient intelligent crowdsourcing model.

I. INTRODUCTION

The notion of IoT has an immense impact on developing the living situation of people through different innovative and revolutionized applications [8][9][11]. Nowadays, the unmanned aerial vehicles (UAVs) are used in different application such as surveillance system, product delivery system, mobile small cell deployment etc. Because of the higher mobility, the UAVs can provide a larger coverage area than any static network infrastructure deployment. Apart from that, the sensory observation or user data is collected by establishing cooperative and non-operative communication in mobile or static sensor network through UAVs[4][5]. Existing research on energy efficient data collection mainly focuses on reducing the energy consumption of the whole network while collecting the environmental sensor data to increase the longevity of the resource constrained WSN or IoT network [6] [7]. Moreover, the existing research on UAV based data collection in a natural calamity mostly focuses on real-time image or video processing to map the affected area without considering the energy efficiency during the flight path of the battery run UAVs [10].

Therefore, in this research, we have focused on solving the energy efficiency in intelligent crowdsourcing for enabling quick response to natural calamity using UAVs and nano data center (NDC). The autonomic crowdsourcing scenario is modeled through the $\epsilon$-greedy Q-learning algorithm and the simulation depicts the efficiency of the proposed model.

II. PROPOSED SYSTEM MODEL

In Figure 1, the nano data centers (NDC) are placed strategically to cover the whole observation area where the NDCs generally collect the sensor observations from the mobile nodes (MN). The mobile nodes are considered as the customer premises equipments (CPE) which produce the crowd data. The NDCs use the base station to send the collected data to the cloud data centers after a certain period of time or when the storage utilization reach to a certain threshold. During the time of any natural calamity, some of the NDCs may become inactive (marked red in Fig. 1) because of the devastating effect of the calamity and therefore the MNs (marked green in Fig. 1) under the inactive NDCs (e.g. D,C) become unable to transmit the sensory observation to the remote cloud data center for the emergency response. Therefore, the UAVs provide the necessary coverage to that infected area so that the environmental data can still become attainable to enhance the efficiency of the rescue operation. The UAVs can communicate with the cellular base station for transmitting the collected sensor data to the cloud data center provided with a seamless and coordinated rescue operation. If the base station becomes unavailable, the UAVs can also transmit the localized crowd data through satellite communication medium and ground station for a quick-response. In our proposed data collection framework, we have considered several scenario which may happen during any natural calamity. For example, during the natural calamity, some of the NDCs become inactive and therefore the MNs are unable transmit the emergency data packet to the cloud. In this case, the UAVs replace the inactive NDCs and collect the observation data from the MNs. However, the UAVs have limited power resource and therefore an energy efficient flight path should be assigned in order to collect the sensory observation from the MNs.

Figure 1: Proposed system model for intelligent crowdsourcing
III. PROPOSED INTELLIGENT ENERGY EFFICIENT CROWDSOURCING

We formulate the energy efficient data collection problem through ”model free” ε greedy Q-learning algorithm [2] [3]. The Q-learning algorithm provides a faster and reliable energy efficient data collection path in terms of coverage area. We consider there are total m number of NDCs strategically placed in the coverage area where k number of NDCs are currently inactive. Therefore the UAVs provides coverage to the state space $S = \{s_1, s_2, ..., s_k\} \cup \{n_1, n_2, ..., n_k\}$ to collect the sensory observations. The agent UAV establishes communication to the MNs which are located to the inactive NDCs and the mobility action is moving from one location state to another. The set of action is $A = \{a_1, a_2, ..., a_k\}$ where k is the number of states. The reward value is the residual energy at each state which is defined as,

$$R = \sum_{l} \left( E_{\text{max}} - E_l(l, p, d, t) \right)$$

In (1), $E_{\text{max}}$ is the maximum energy of the UAV, $E_l$ is the energy consumption where l is the received packet length from ith MN, $p$ is the transmission power, $d$ is the data rate and t is the flight time. The priority of any particular state is defined thorough the weight factor w as in,

$$w = \left[ \frac{i - i_{\text{min}}}{i_{\text{max}} - i_{\text{min}}} \times (h_{\text{max}} - h_{\text{min}}) + h_{\text{min}} \right]$$

where i is the number of active MN for crowdsourcing, $i_{\text{min}}$ is the minimum number of MNs, $i_{\text{max}}$ is the maximum number of MNs connected to each UAV. $h = [1, 2, 3]^T$ is the priority scale vector where value 1 is the least priority state and 3 the high priority state. The agent UAV explores from one state to another until it reaches the goal state and thus converges. At time t, for each action $a_t$ in state $s_t$ the Q-value is updated as in,

$$Q(s_t, a_t) = (1 - \alpha_s)Q(s_t, a_t) + \alpha_s (R + \beta \max\{Q(s_{t+1}, a_{t+1})\})$$

(2) $Q(s_t, a_t)$ is the Q-value of the current state and $Q(s_{t+1}, a_{t+1})$ is the Q-value of the next state, $\alpha = [0, 1]$ is the learning rate, $\beta$ is the discount factor. The Q-value is updated with probability $\epsilon$ and stored in the Q matrix which acts as the memory of the agent UAV.

IV. PERFORMANCE EVALUATION

Table I represents the simulation parameter of the proposed algorithm. We choose the initial state of the UAV agent 2 and the goal state is set to 86. The flight time t depends on the cumulative distance between that the agent UAV transverses. In Fig.2, the algorithm converged to the shortest path after 3000 episodes and the number of exploitation is 44080 and the number of exploration is 31027. The UAV transversed 10 states to reach the goal state. The cumulative reward gain of UAV is around 6.029954e+002 at the time of convergence.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>MN</td>
<td>500</td>
</tr>
<tr>
<td>UAV</td>
<td>1</td>
</tr>
<tr>
<td>NDC</td>
<td>100</td>
</tr>
<tr>
<td>Penalty, $\epsilon$</td>
<td>0.9</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.90</td>
</tr>
<tr>
<td>$E_{\text{max}}$</td>
<td>22.2V</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.1V</td>
</tr>
<tr>
<td>$t$</td>
<td>10000 bit</td>
</tr>
</tbody>
</table>

(a) Steps vs. Episodes  
(b) Reward vs. Episodes

Figure 2: Q-learning performance

V. CONCLUSION

In this paper we have proposed the energy efficient intelligent crowdsourcing during natural calamity through UAVs. In future, we will extend the research by implementing the prototype considering more diverse parameters.

REFERENCES


Data Curation Layer for Wellness Platforms

Hafiz Syed Muhammad Bilal
Department of computer science and engineering,
Kyung Hee University,
Yongin, South Korea,
bilalrizvi@oslab.khu.ac.kr

Sunyoung Lee
Department of computer science and engineering,
Kyung Hee University,
Yongin, South Korea,
sylee@oslab.khu.ac.kr

Abstract—In this paper, we present our third iteration of Data Curation Layer (DCL) for wellness platforms. DCL performs device-independent accumulation of raw sensory data from multimodal data sources in real-time. Furthermore, it curates this data with associated user context over user life-log. DCL is equipped to monitor the lifelog instances for situations of interventions. For the permanent storage of large volume of raw sensory data, DCL uses big data platform. Persisted data is aligned to be reused for Analytics and visualization. DCL encapsulates all the computational complexity at the cloud keeping the client as thin as possible.

Keywords—Data curation; Lifelog; Wellness platform; Data acquisition; Big data

I. INTRODUCTION

With the advent of smart devices, an opportunity has emerged for the healthcare providers and biomedical researchers to enable people for taking care of their health and wellness, by providing them timely, ubiquitous, and personalized support [1]. As a result, the penetration of fitness wearables with smartphone applications and systems supporting health and wellness have taken the market by storm [2]. Despite this enormous effort by the industry and research, most of the current solutions are single device focused on the limited scope and lack interoperability and performance [3]. Considering the limitations of existing efforts as an opportunity, we have proposed and implemented a comprehensive health and wellness platform called Mining Minds [4].

The core of the mining minds platform lies in its foundations built for accumulation, persistence, and monitoring of raw sensory data. These responsibilities are undertaken by the novel layer of Data Curation (DCL). This layer is designed for high-performance acquisition of a variety of user-based raw sensory data in real-time; furthermore, persistence as well as monitoring of context of this sensory data. Unlike other healthcare and wellness platforms where sensory data acquisition is device specific, DCL facilitates mining minds to communicate with a variety of multimodal data sources registered to a particular user for the sensory data acquisition making it compatible with an IoT-based health and wellness environment. Furthermore, DCL curates the sensory data in a user-based timeline model called user life-log. Dissimilar to modern healthcare and wellness platforms, this log provides continuous monitoring abilities to DCL, so that, it can identify any anomalies where a user might require automatic support or assistance from the platform. Moreover, DCL is equipped with big data support for permanent persistence of user’s sensory data. This data is further used by other layers of mining minds platform for visualization and predictive analytics.

II. METHODOLOGY

As illustrated in fig 1, DCL consists of two primary components, i.e., Sensory Data Processing and Curation, and Non-volatile Sensory Data Persistence. Within the former, sensory data acquisition supports the acquisition and synchronization of raw sensory data obtained from multimodal data sources, both in real-time (active) and offline (passive) manner. Acquired sensory data is firstly synchronized based upon the user and the timestamp of data generation, and secondly sent to ICL [4] for context identification. In response, lifelog mapping and representation component receives the identified context from ICL and curate it by mapping the context instances to a time-based log registering the detected human activities and behaviors. This log is termed as user LifeLog or simply Lifelog and persisted in the Intermediate Database for the share-ability with participating layers and applications. The stream of lifelog instances is analyzed by a monitoring component called Lifelog Monitor (LLM). It is responsible for performing time-based monitoring of different user attributes and variables, hosted in the lifelog, cross-linked with their user profiles. Furthermore, LLM supports trigger-based mechanisms to notify SCL [4] for the occurrence of an abnormal or special event associated with a given user. This mechanism is the basis for push-based recommendation generation and notification to the user by the mining minds platform. Abnormal events monitored by LLM represent risky or unhealthy behaviors and are here defined as Situation-events or situations in general. These situations are described through various constraints (e.g., age, gender, medical conditions) and monitor-able variables (e.g., the intensity of a particular activity and its duration). Situation events can be generated both statically at design time and dynamically at run-time by a domain expert via KCL [4].

DCL's Non-volatile Sensory Data Persistence is responsible for providing permanent and distributed big data persistence to the raw sensory data. Non-volatile sensory data persistence provides mechanisms to access the persisted data as a response to an active or a passive request generated by participating layers and applications. For online requests by SL [4] for visualization and analytics on incoming raw sensory data, the subcomponent of Active Data Reader is utilized. For the training of models used for generation of rules by KCL,
Passive Data Reader subcomponent is used. LLM further utilizes these rules for its trigger-based mechanism. To create periodic backups of lifelog, the subcomponents of Lifelog Sync is used. This method provides permanent storage to the user’s lifelog which can be utilized in future for human behavior analysis.

III. CONCLUSION

In this paper, we presented the core layer of mining minds platform called data curation (DCL). This layer is responsible for the acquisition of raw-sensory data in real-time. Its data source independent nature makes it more scalable and IoT compatible. DCL enables the multimodal sensory data sources to send data directly via REST service. Its logical clock synchronization enables the platform layers to identify the right context of the user at an instance of time. DCL maps the context derived by ICL from the raw-sensory data over a time-based user life-log. Furthermore, it monitors this lifelog of registered users for the detection of situations in which a push from the platform needs to be invoked to alert the user. This monitoring can integrate static, dynamic, and complex situations created by the data-driven approach of KCL. DCL incorporates multi-level abstraction on data depending upon its usage and persistence. Frequently required user lifelog and profiles data is persisted in an intermediate database hosted over an RDBMS; whereas, the historic and raw sensory data is persisted in a non-volatile storage provided by big data technologies. Active and passive data read of DCL facilitates the SL for analytics and visualization, and KCL for its model training to generate data-driven knowledge respectively.

ACKNOWLEDGMENT

This work was supported by the Industrial Core Technology Development Program (10049079, Develop of mining core technology exploiting personal big data) funded by the Ministry of Trade, Industry and Energy (MOTIE, Korea).

REFERENCES

Abstract—Smartphone contains many confidential information about users like passwords, bank number and many more. Therefore, the need for reliable security for the smartphone is very important. In this paper, it explained about the design and development of a system based on the behavior biometrics authentication of an individual on the rotation touch behavior. The movement of the fingers is studied by using the statistical calculation like standard deviation. Besides that, FAR and FRR are very important for the process of the behavior touch operation authentication.

Keywords—Behavior biometrics authentication; rotation touch behavior; FRR; FAR; standard deviation

I. INTRODUCTION

A smartphone is very important to the user as it can be a computer, a phone, a data base a camera and a locator. In other words, a smartphone is a very personal device that the users save their personal data, passwords, health information and much more critical items [1]. The most advance method in authentication of smartphone is biometric authentication. Biometric authentication can be divided into two which is physical and behavior authentication [2]. Fig. 1 shows the different types of biometric authentication.

Nowadays, traditional authentication methods using passwords and identity documents are not enough in order to protect the users and easily to be hacked by hacker. The traditional authentication methods are easily lost, forgotten, guessed, shared or stolen. The problem in the traditional authentication methods is solved by using biometrics authentication which is proven in the reliability and security [3][4].

In recent years, biometric authentication has been used for replacing the traditional authentication methods for smartphone [5][6][7]. Generally, biometric authentication is consist of two different modes which is verification and identification. Verification is the system where the user claim to be already in the system and the user biometric data is compared to the user’s data that stored in the systems.

Meanwhile, for identification, the identity of the user is unknown but biometrics of user is compared to the existing record at anywhere in the database. The identification process is more challenging compare to verification. This paper describes the behavior biometrics authentication based on touchscreen and using verification mode.

II. BACKGROUND OF REFERENCE THRESHOLD, FRR AND FAR

A. Reference Threshold

Reference Threshold is the defined as a value that accepted or rejected users is determined by using biometric authentication. The value of reference threshold is very sensitive in perception of accuracy as authentication depends on reference threshold chosen. Training of the system, hereby touchscreen of finger movement, is done to find suitable value of the reference threshold [8].

For this Research, the value of the reference threshold is determine by the value of standard deviation. On the verification process we will find the mean and the standard deviation of each users touch operation.

B. False Rejection Rate (FRR)

FRR is defined as rejected user being rejected because of the criteria of reference threshold is not fulfilled and the accepted user is lying in the range of the rejected user.

C. False Acceptance Rate (FAR)

Meanwhile, FAR is defined as rejected user being authenticated as accepted user because the criteria of reference threshold is fulfilled and the rejected user is lying in the range of accepted user.

III. PROPOSED METHODS

The proposed method of authentication system is shown in Fig. 2. The systems will have two processes, verification (enrollment) and authentication. Firstly, the user will do an enrollment to the system and keep in the database of the accepted user. The next process is verification of the behavior biometrics authentication which is rotation touch behavior. Upon completion of the enrollment and verification process, the user able to use his/her behavior biometrics authentication.

A. Enrollment and Verification Process

First the user must enroll him/herself by creating a new profile and keyed in a few information about themselves. Fig. 3 shows the verification process for rotation touch behavior and then the information is keep in the database of the system.
Fig. 2. Proposed Authentication System

Fig. 3. Verification process for rotation

Fig. 4. Reference Threshold for user B

Table 4 shows the data attained from the process of verification for user B. The reference threshold is obtained for the behavior biometrics authentication of user B. The black linear line is the mean of the rotation duration for user B.

B. Authentication Protocol

In Authentication Protocol, the user has to type in the Profile ID and proceed to authentication process. Then, the user has to use his/her behavior biometrics as the user enrolled and in this case it is using the rotation touch behavior. The result of the action then will be compared to the user database where the reference threshold biometrics plays a vital role in order to determine he/she is the enrollment user. The display will show the access granted or access denied respectively upon the user completed the authentication protocol.

Table I shows the average of rotation duration for each users. From Table I, we can see that user A have the longest average of rotation duration compare to the other users. This is probably user A tend to rotate his/her finger slowly. The standard deviation for user B is obtained from the data of Fig. 4 and then the results of user B is tabulated in Table I and Table II. The pattern of the user B on the verification process is linear at the first 25 times. However, maybe the process of verification is too long, after that the pattern become scattered.

### Table I. Average of rotation duration

<table>
<thead>
<tr>
<th>User</th>
<th>Rotation duration (secs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>5.49 ± 2.50</td>
</tr>
<tr>
<td>B</td>
<td>3.76 ± 1.46</td>
</tr>
<tr>
<td>C</td>
<td>3.72 ± 2.19</td>
</tr>
<tr>
<td>D</td>
<td>2.83 ± 0.45</td>
</tr>
</tbody>
</table>

### Table II. FRR and FAR experiment data of rotation

<table>
<thead>
<tr>
<th>User</th>
<th>Rejected</th>
<th>FRR</th>
<th>Accepted</th>
<th>FAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>3 of 20</td>
<td>15%</td>
<td>2 of 20</td>
<td>10%</td>
</tr>
<tr>
<td>B</td>
<td>6 of 20</td>
<td>30%</td>
<td>1 of 20</td>
<td>5%</td>
</tr>
<tr>
<td>C</td>
<td>4 of 20</td>
<td>20%</td>
<td>2 of 20</td>
<td>10%</td>
</tr>
<tr>
<td>D</td>
<td>10 of 20</td>
<td>50%</td>
<td>0 of 20</td>
<td>0%</td>
</tr>
<tr>
<td>Average</td>
<td>29%</td>
<td>6%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

User C have almost the same rotation duration with user B but the reference threshold is wider and as the result shows in Table II user C have lower FRR rate compare to user B. User D have the fastest rotation duration and the shallowest reference threshold but the results is half of the authentication process falsely reject user D. This maybe because it’s strict and rotation touch behavior is hard to do.

IV. CONCLUSION

This paper shows a new method of behavior biometrics authentication system based on rotation touch behavior for smartphone which has been implemented on the smartphone with Android OS. From our opinion, the averages of FRR and FAR is high and not good. The user have to authenticate themselves at least 5 times if there is any false rejection. And it took so much time to just authenticate. Therefore, we hope in the future to add more variables on the rotation touch behavior. For examples, rotation speed, finger positioning and do clock-wise rotation and vice-versa. This maybe would improve the authentication system and reduce the FRR and Far percentages.

REFERENCES


Local and Global map interface design for flying drone

Sanghyeong Yu  
Graduate Program in Cognitive Science, Yonsei University, Seoul, Republic of Korea  
shyu.ux@gmail.com

Jongwoo Jeon  
Graduate Program in Cognitive Science, Yonsei University, Seoul, Republic of Korea  
mashi79513@gmail.com

Kwangsu Cho  
Graduate School of Information; Graduate Program in Cognitive Science, Yonsei University, Seoul, Republic of Korea  
kwangsu.cho@gmail.com

Abstract— To fly a drone safely, its pilot needs to maintain high situation awareness of its flight space. One of the important ways to improve the flight space awareness is to integrate both the global and the local navigation map a drone provides. However, the drone pilot often has to use the inconsistent reference frames or perspectives between the maps. In other words, the global navigation map tends to display space information in the third person perspective, whereas the local map tends to use the first person view through the drone camera. This inconsistency results in the so-called inconsistent perspective problem, which requires the pilot to use mental rotation to align the different perspectives. In addition, the two maps use different dimensionalities (2D vs. 3D) that may aggravate the pilot’s cognitive load of the mental rotation. Therefore, this study aims to figure out the optimal way of reducing the pilot’s mental rotation when using the maps in terms of the angle of perspective difference (0˚, 90˚, 180˚, 270˚) and the map dimensional matches (3D-3D vs. 3D-2D). The results show that the pilot's flight space awareness improves when the perspective differences are smaller and also when the dimensionalities between the two maps match.

Keywords— drone; global map; local map; spatial situation awareness; mental rotation; reference frame; dimensionality; navigation

I. INTRODUCTION

Recently, civil drone accident is occurred consistently despite it is equipped the automatic navigation system[1] to prevent developing the cognitive load of pilot. According to UAV(unmanned area vehicle) accident data of military, due to civil drone has not sufficient accident data, 33% of 70% of control accident is occurred by human mishaps[2]. As aforementioned, we assumed that civil drone accident is happened more than military UAV accident.

One of the critical drone accident cause is the deficiency of pilot’s situation awareness of its flight space. The situation awareness is similar concept to geographical awareness and navigational awareness. It helps pilot to developing mental model that enable to find out the location and surroundings of its flight[2-5].

Using the external piloting which pilot have to watch it through their eye, the civil drone pilot tends to have difficult with understanding the flight space[6-9]. Civil drone pilot usually use the ego-centric reference frame or first person view perspective. But, pilot also needs third person view like allocentric reference frame that represented drone’s view or world-centered reference frame to fly drone safely.

To solve this problem, current drone controller offers pilot to understand the flight space easily using local and global map together.

Taking up the most space in drone controller, local map shows the part of surrounding the drone with detailed information. The local map, which field forward view screen of controller, is usually consisted of video from the drone’s camera. So it is possible for pilot to pretend that they fly drone through internal piloting which pilot should seat the inside the drone[10]. Recently, it is developed that pilot can fly the drone through the immersive head-up display which show the video from drone’s camera(emp. [11-13]). But it is restricted that pilot’s viewing angle through the video from drone’s camera. The range of viewing angle of civil drone is from 90˚ to 140˚. Considering the range of viewing angle of general setting for piloting is from 180˚ to 270˚, civil drone pilot undergoes a difficult time to understand the overall flight space.

On the other hand, the global map is appropriate for drone controller interface which solve the problem of the local map. Located in below the local map on controller, the global map is offered widen geographical information with downscaled. Also, It is suitable to show broaden space surrounding the drone with third person view, like global positioning system(GPS). According to [14], the global cannot show the detailed information as size is small. But, it can show the approximate location and direction from the broaden space.

Thus, using the local and global map together may be efficient to drone navigation as there are complementary that local map is offer detailed information of small space and global map is show the approximate information of broaden space[14], [15].

October 27-28, 2016
Beyond AlphaGo
PACS2016
However, against the original intention, it aggravates the maintain high situation awareness of its flight space to using the local and global map together. Since, pilot have to integrate the space information with referenced the two maps in real-time. It causes the difficulty to pilot that integrating the inconsistent reference frame between two maps. The reason is that the local map is a rotation map which north-up direction is changed continuously along the field forward view, the global map is a fixed map which north-up direction is unchangeable[16]. Therefore, pilot is influenced by the angle of perspective difference between local and global map.

In addition, whether map is three-dimensionality or two-dimensionality is affect the pilot’s situation awareness of its flight space. According to preceding research, three-dimensionality is more effective to navigation as it is the representation that compatible with real space, and the follow the proximity compatibility principle[17-19]. Otherwise, the mental rotation speed of two-dimensional map is faster than three-dimensional map. Also, two-dimensional map is more beneficial for pilot to process the two-dimensional information like air speed control. Accordingly, the advantage of map dimensionality is different by which context of flight is offered.

Until the present, the research and development is focused on not what factor of local and global map is enable navigation more efficient, but why we use local and global map together in navigation. Therefore, this study aims to figure out the optimal way of reducing the pilot’s mental rotation when using the maps in terms of the angle of perspective difference.

II. SOLUTION

This study aims to figure out the optimal way of reducing the pilot’s mental rotation when using the maps in terms of the angle of perspective difference (0˚, 90˚, 180˚, 270˚) and the map dimensional matches (3D-3D vs. 3D-2D). To experiment, we measured the accuracy which means whether they select correct flight direction or not and reaction time that means whether they select flight direction easily or not.

Following the preceding research, we assumed that the subject will spend more time to select the flight direction if angle of perspective difference is large. Also, the subject will spend short time to select the flight direction if two maps are consistent with reference frame.

III. CONCLUSION

In practical perspective, this research finds out which global map is easy to retrieval the spatial information when three-dimensional field forward view is used together. Furthermore, it will contribute to develop the navigation design for fly drone safely.

In theoretical perspective, it will contribute to understand how pilot develop the situation awareness of its flight space through using both local and global map. Therefore, this outcome will help to reduce the drone accident.

REFERENCES

Abstract—The focus of healthcare and wellness technologies has shifted towards personal vital signs devices. From health applications over the smartphones, the technology has evolved to be enclosed into devices with much smaller form-vector like fitness bands and smartwatches. The novelty of these devices is the accumulation of sensory data as its users go by their daily life routine; consequently, empowering themselves in self-management as well as enhancing traditional wellness and healthcare procedures. Although substantial domain-specific contributions have been made; however, due to device specific implementations, these contributions fail to execute as independent data accumulation platforms. In this paper, we present the third iteration of Mining Minds, open source health and wellness platform.

Keywords—Health and Wellness platform, Data curation, Knowledge engineering, wearable devices, Human behavior, Digital health, Wearable sensors, Big data, Cloud computing, Context-awareness, Knowledge bases

I. INTRODUCTION

Recently, a shift has occurred in the theory of handling healthcare and its supporting systems. This change has made a substantial impact on the intention of healthcare models. Rather than late disease management and cure, these models are focusing on preventive, personalized health. Therefore, service providers are pushing forward for wellness based models and conducting researches to investigate their effectiveness. Advanced studies in biomedical healthcare have shown that the most common diseases are partly caused by the poor lifestyle that people maintain in their daily routine. Unhealthy and fast-food diets, use of tobacco, sedentary routines with the lack of exercise are among the potential contributors to develop illnesses and also limit the effectiveness of medical treatments [1] [2] [3]. In spite of this enormous effort by the industry and research, most of the current solutions are single device focused with limited scope and lack of interoperability and performance [4]. Considering the limitations of existing efforts as an opportunity, we have proposed and implemented a comprehensive health and wellness platform called Mining Minds [5].

Mining Minds is built on the core ideas of digital health and wellness paradigms to enable the provisioning of personalized support [5]. It is built on the utilization of prominent digital technologies ranging from real-time raw sensory data acquisition, big data, cloud computing to wearable and Internet of things (IoT), as well as modern concepts and methods such as context-awareness, knowledge basis with analytics, to holistic and continuous investigation on people’s lifestyle and provide a variety of smart coaching and support services.

II. METHODOLOGY

Mining Minds is a state-of-the-art wellness platform that is designed to be a layered architecture as illustrated in the fig 1. Each layer pertains to the abstraction on raw sensory data acquired from multimodal data sources. Starting from top, Service Curation Layer (SCL) is responsible for curating recommendation based services based on several different user scenarios; Knowledge Curation Layer (KCL) is responsible for curating data- and expert-driven knowledge that is further extracted as rules which play decisive role in user lifelog monitoring and recommendation generation; Information Curation Layer (ICL) is responsible for context identification of the user based on the raw sensory data; and Data Curation Layer (DCL) is the foundation of Mining Minds platform, responsible for not only real-time data acquisition but also its non-volatile persistence. Furthermore, the user context identified by ICL is continuously monitored by DCL based on the rules provided by KCL for situation detection in which a user might require assistance in the form of recommendations via SCL.
Fig. 1. Mining Minds abstract architecture

Mining Minds platform is a hybrid cloud implementation with raw-sensory data and its processing components hosted on a private cloud; however, services and other layers are deployed on MS. Azure public cloud platform. For the sharability of this implementation, we intend on releasing Mining Minds as open source platform. This intention will enable health and wellness researchers to collaborate and extend our implementation to further possibilities and usages.

III. CONCLUSION

This work presents Mining Minds, a novel digital framework for personalized healthcare and wellness support. The framework has been designed by taking into consideration the requirements of health and wellness systems. The unique architecture of Mining Minds platform is built to support the necessary functionality that enables curation of data, information, knowledge, and services for personalized health and wellness support.

REFERENCES


SaKEM: A Semi-automatic Knowledge Engineering Methodology for Building Rule-based Knowledgebase

Maqbool Ali, Maqbool Hussain, Sungyoung Lee*
Department of Computer Science and Engineering
Kyung Hee University
Yongin, South Korea
{maqbool.ali, maqbool.hussain, sylee}@oslab.khu.ac.kr

Byeong Ho Kang
School of Engineering and ICT
University of Tasmania
Hobart, Australia
byeong.kang@utas.edu.au

Abstract—Knowledge engineering is one of the key research area to build knowledgebase for providing solutions to real-world problems. Due to rapidly increase of data growth rate, it is almost impossible to extract hidden knowledge with manual approach. Moreover, a number of methodologies have been proposed that focus on some specific aspect of data mining process rather than end-to-end knowledge engineering methodology. Keeping in view these facts, a Semi-automatic Knowledge Engineering Methodology (SaKEM) is proposed that covers all major stages that are involved in Knowledge Discovery in Databases (KDD) process. For realization of SaKEM, a toolset called Data Driven Knowledge Acquisition Tool (DDKAT) is developed. The proposed methodology is designed for Mining Minds project but it can be utilized by other service-enabled platforms as well.

Keywords—features selection; data preprocessing; decision trees; model translation; production rules; knowledge acquisition

I. INTRODUCTION

Knowledge is a key to compete and succeed in every field of life, which can be discovered from heterogeneous sources by various knowledge discovery approaches. In current arena, almost every organization need decision support system that use discovery knowledge techniques for making better decisions. Normally knowledge is acquired either by expert-driven, where the expert heuristics are used or data-driven, where state-of-the-art data mining methods are applied. Due to rapidly increase of data growth rate, it is almost impossible to extract hidden knowledge with manual approach [1]. This paper proposed a Semi-automatic Knowledge Engineering Methodology (SaKEM) for building a knowledgebase using data-driven as well as expert-driven approaches. The proposed methodology consists of five important phases namely data selection, data preprocessing, model learning, model translation, and rule conformance that are realized through Data Driven Knowledge Acquisition Tool (DDKAT).

The SaKEM is designed for Knowledge Curation Layer of Mining Minds (MM) project. The MM is a platform which will take benefit from the technology of big data with respect to variety as well as volume, mapping of life events through sensory environment and reasoning and prediction to process the real-time data for providing personalized services [2]. The multilayer architecture of MM framework is depicted in Fig. 1, where SaKEM is elaborated. Each layer of MM framework has specific tasks such as the Data Curation Layer is responsible for obtaining data from Multimodal Data Sources, Information Curation Layer for describing the user context as well as behavior, Knowledge Curation Layer for developing the health and wellness knowledge, Service Curation Layer for creating health and wellness support services, and Supporting Layer for providing security as well as data visualization facility [3].

II. METHOD

A. Data Selection

The first step in any decision support system is to understand the application domain and then to identify application goal, objectives, causative factors, and their associations. All these factors help data understanding phase. Based on expertise, domain expert selects suitable parameters from available features list. In addition to that, feature selection measures also assist the domain expert in selecting informative features for decision making [4].
B. Data Preprocessing

Once a selected feature-set is obtained from a dataset, then data pre-processing is required that can play major role in improving the quality of data. This phase performs the basic data preprocessing tasks such as (1) identification and replacing of missing values with attribute’s mean/mode mechanism, (2) detecting and replacing outlier values with Interquartile technique and attribute’s mean mechanism respectively, and (3) data discretization with Equal-Width Binning and Equal-Frequency Binning methods.

C. Model Learning

The aims of this phase is to learn and build the classification model, called decision trees (DTs). In order to understand the knowledge structures, most of the health-care experts are interested in knowledge visualization or its representation [5]. The DTs are one of the popular data classification methods that represents the inductive knowledge. In this phase, DTs are built using BFTree, J48, J48graft, RandomTree, REPTree, and SimpleCart classification algorithms.

D. Model Translation

The decision trees are non-executable format and there is need to extract rules and convert them into executable format i.e. production rules. This is not an easy task and very limited information is found to translate the decision tree into production rule set [6]. To achieve this goal, this phase performs model trimming, XML conversion, XML parsing, and rules conversion tasks.

E. Rule Conformance

The aims of this phase is to build the trusted knowledgebase. To achieve this goal, all production rules are shown to domain expert through expert-driven interface, where expert conforms them one-by-one and stores them into knowledgebase.

III. RESULTS: CASE STUDY FOR USERS PROFILE AND LIFE-LOG MINING MINDS DATASET

We have applied the proposed methodology in digital health and wellness paradigms that monitors the users’ daily life activities to provide healthy habits recommendations [2]. Mining Minds Users Profile and Life-log Dataset is selected to realize the SaKEM, where partial outcome of this methodology is illustrated in Table I.

<table>
<thead>
<tr>
<th>Rule</th>
<th>Production Rules</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>IF (SituationCategory ≠ LyingDown OR Sitting) THEN Recommendation = Sitting</td>
</tr>
<tr>
<td>2</td>
<td>IF (SituationCategory = LyingDown OR Sitting AND RiskFactor ≠ Normal) THEN Recommendation = Stretching</td>
</tr>
<tr>
<td>3</td>
<td>IF (SituationCategory = LyingDown OR Sitting AND RiskFactor = Normal AND Age ≥ 32) THEN Recommendation = Walking</td>
</tr>
</tbody>
</table>

IV. CONCLUSION

This paper demonstrated the end-to-end knowledge engineering methodology for building a knowledgebase using data-driven as well as expert-driven approaches. The methodology is demonstrated with users’ profile and life-log dataset and outcome of SaKEM is used in real project of Mining Minds.

ACKNOWLEDGMENT

This work was supported by the Industrial Core Technology Development Program (10049079, Develop of mining core technology exploiting personal big data) funded by the Ministry of Trade, Industry and Energy (MOTIE, Korea).

REFERENCES

X-UDeKAM: An Intelligent Method for Acquiring Declarative Structured Knowledge using Chatterbot

Maqbool Ali, Jamil Hussain, Sungyoung Lee*
Department of Computer Science and Engineering
Kyung Hee University
Yongin, South Korea
{maqbool.ali, jamil, sylee}@oslab.khu.ac.kr

Byeong Ho Kang
School of Engineering and ICT
University of Tasmania
Hobart, Australia
byeong.kang@utas.edu.au

Abstract—Declarative knowledge has a key role in health-care domain, which is hidden in multiple clinical-related knowledge resources. For acquiring and constructing the declarative structured knowledge from unstructured knowledge resources, UDeKAM methodology was proposed that lacks the intelligent human dialoguing mechanism for validating and filtering the important concept relations. This paper is the extension of our previous work and presents an extended version of UDeKAM, called X-UDeKAM, that includes Chatterbot along with text preprocessing as well as controlled natural language processing techniques. For realization of X-UDeKAM, a diabetes scenario is explained through example.

Keywords—declarative knowledge; text processing; controlled natural language, chatterbot; unstructured knowledge;

I. INTRODUCTION

Knowledge is a key to compete that solves real-world problems and provides help in decision making. Declarative knowledge is a category of knowledge and expressed in the form of unstructured sentence. In health-care domain, there exists plenty of declarative knowledge which is very critical for quality of health management. Acquiring knowledge from unstructured knowledge resources is one of the key research areas, where multiple systems exist that have been proposed for particular applications without support of controlled natural language and intelligent human dialoguing mechanism [1-4]. Keeping in view these facts, an extended version of UDeKAM [1], called X-UDeKAM is proposed. The architecture of the proposed methodology is depicted in Fig. 1, where text preprocessing, Chatterbot, and model construction using controlled natural language phases are elaborated.

II. METHOD

A. Text Preprocessing

The first step is to collect the domain resources i.e. diabetes documents. After collection process, this phase applies basic Natural Language Processing (NLP) techniques [4] and extracts the concept relations. For example, following are the two relations of word ‘symptom’ that are extracted from diabetes standard guidelines:

1. symptom/feeling/numesthesia/unpleasant_person/negative_stimulus/hurt
2. symptom/blood_disease

B. Chatterbot

A chatterbot is an intelligent computer program, which helps to simulate human conversation via textual methods [5]. Once concept relations are obtained, then there is a need for validation and filtration of important relations. This phase utilized intelligent dialoguing mechanism with the help of Artificial Intelligence Markup Language (AIML). For example, for validation of “blood_disease” relation, here is a very simple conversion between human and virtual agent is:

<topic name="diabetes-symptoms">
   <category>
      <pattern>What are symptoms of diabetes patient?</pattern>
      <template>Possible symptoms are <bot name="symptom"/>.</template>
   </category>
   <category>
      <pattern>Is blood_disease a symptom of diabetes?</pattern>
      <template>Yes.</template>
   </category>
</topic>

C. Model Construction

After validation of concept relations, there is a need to construct domain model, which is consist of declarative knowledge. This phase used Attempto Controlled English (ACE) and OWL for constructing declarative structured knowledge [6, 7], which is partially shown in Fig. 2.
III. CONCLUSION

This paper demonstrated an extended version of declarative structured knowledge acquisition methodology with support of intelligent human dialoguing mechanism.

ACKNOWLEDGMENT

This work was supported by the Industrial Core Technology Development Program (10049079, Develop of mining core technology exploiting personal big data) funded by the Ministry of Trade, Industry and Energy (MOTIE, Korea).

REFERENCES


Abstract—To create intelligent systems for use in personal robots, the integration of information from various sensors to conceptualize knowledge and understand complex events are crucial. In this paper, we attempt to build a cognitive model which integrates various modalities of observed people (location, object recognition, reading emotions and activity recognition) to create a plan library which is further used to build a plan recognition model. We present an improved method of plan recognition by developing an automated plan library builder which addresses the limitations of existing predefined plan libraries. This provides a more robust system, more suited to functioning in real-life situations where the infinite number of actions can be managed. Experiments were undertaken in real time, in a real-life home environment setting with a specially built turtlebot and human subjects. Our proposed system consists of two parts, first the human modality collecting module, second the automatic plan library builder and plan recognition module by the Deep Concept Hierarchy (DCH) model. The results from our model show promising performance in the designed plan recognition. This demonstrates the integration of multimodal information and conceptualization of knowledge for an interactive system which efficiently infers and responds to the goals and plans of observed people.

Keywords—Cognitive Robotics, Machine Learning, Plan Recognition

I. INTRODUCTION

The development of AI is entering a new phase where it has progressed from expert systems to building autonomous robots to assist people. As a result, many are now expecting the agents to be smart enough to support them with more complex tasks, understand emotions and behaviors, learning from their lives and even protecting them from danger. Therefore, in this paper, we propose a new approach for designing a cognitive model for intelligent systems, which connects concept learning with plan recognition.

II. RELATED WORK

For a robot to assist people, it is essential for the robot to have a conceptual knowledge to recognize the current situation and decide the best action to respond. Much of the current research on building knowledgeable concepts has approached through probabilistic inference [1], and applications to visual-linguistics construction [2]. Moreover, for achieving the ability of situation awareness, [3-4] has reported significant results for recognizing other agent’s situation or goal from plan recognition. However there exists a strong constraint in plan recognition algorithms which needs a complete set of rules to recognize situation or goal of other agent. Therefore by employing the methodology of probabilistic concept construction and online learning [5] to plan recognition, we present a cognitive architecture that autonomously builds and recognizes plans of human for the robot to provide personal caring services.

III. METHODOLOGY

To build an autonomous personal caring robot, various modules are needed. The basic modules necessary for the system to execute are listed below:

- SLAM
- Human detector
- Human position estimator
- Human follower
- Modality(whom, object, place, activity) recognizer
- Plan library builder and recognizer

In this paper we mainly discuss about the plan library builder and recognizer.

A. Plan Library Building and Recognizing

The proposed DCH model was originally used to learn concepts by automatically constructing knowledge from visual-linguistic information [2]. However, modifications can be made to design a DCH model which captures the characteristics of a plan library to build a plan library with enhanced abilities. The design of our DCH model is elaborated below.

First, the hyperedges in the second layer of the DCH model can be interpreted as the basic actions in the plan library. The observation set(φ) in Fig.1(top) includes the recognized outputs, such as who, place, emotion, activity and the time of occurrence. This set of hyperedges becomes the primitives in the plan library used in the plan recognizer:

\[
\phi \in A \text{ when } V(\phi) > r
\]

\[
\Delta E = \Delta E \cup \{\phi\} \text{ and } \phi = \bigcup_{m=1}^{k} \phi_m
\]

\[
P(\phi) = \prod_{m=1}^{k} P(\phi_m)
\]
Where $V(o)$ is the vertex value and $r$ is set to 30, $\Delta E$ indicates the new hyperedge set and $e$ is the created hyperedge. $P(o)$ is the total probability of each component in $o$.

With the created primitive, we learn the relationship between the primitive by the graphical Monte Carlo method.

$$C' = \arg \max_{C} P(G_{i} | O) = \arg \max_{C} P(O | G_{i}) P(G_{i-1})$$

By estimating the $P(G_{i} | O)$ to be maximum while observing observations, the library builds up to find the optimal relations of the primitives.

To recognize the plan, the recognizer infers the maximal probability of the relation from the observation.

IV. EXPERIMENT

We examined the performance of the plan library recognizer by letting three participant act a given sequence of tasks. The plan recognizer should recognize the same goal/scenario of every participant. The sequence of tasks from each participant should be recognized and by the plan library to infer the same goal/scenario. The sequence of the scenario is as follows; sit on the couch $\rightarrow$ go to the kitchen to eat dinner $\rightarrow$ go to the couch and play with the laptop $\rightarrow$ lay down on sofa $\rightarrow$ go to room and play $\rightarrow$ walk around the living room $\rightarrow$ sit down on the sofa. The duration of each action was not restricted. The total time of test sequences collected from all participants was 5461 seconds and the ground truth sequence of the given scenario was labeled by hand. The accuracy of the recognized goals of participants was evaluated by calculating the mean-squared error of the ground truth label and the labels of goals derived from the plan library after observing the participants. The accuracies are shown in Table 1, where the average accuracy was 89%.

<table>
<thead>
<tr>
<th>Participant</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observation Size</td>
<td>1660</td>
<td>1753</td>
<td>2048</td>
</tr>
<tr>
<td>Accuracy (%)</td>
<td>91%</td>
<td>85%</td>
<td>90%</td>
</tr>
</tbody>
</table>

These results show that a plan library build by DCH model could be used to recognize observed sequences to infer the goals of humans in the home environment.

V. DISCUSSION

We have proposed a new approach of plan recognition with a home environment robot system to present the potential of adapting an AI agent into the real-world. Our DCH model demonstrates high performance in building new concepts (primitives) with the relations between them to build a plan library for plan recognition. This occurs in an incremental, unsupervised manner, where observations in real-time are collected, to form the primitives necessary to build the concept

![Figure 1. Architecture of Deep Concept Hierarchy](image)
Abstract—It is important for a personal or service robot to be
learned autonomously from the real world. This paper proposes
an approach to developing an autonomous schedule learning and
acting robot platform in dynamic home situations. Experimental
setup of a pseudo-real home environment with a scenario and
robot platform is described. Ongoing results including visual
scene understanding and event prediction are presented.

Keywords—home robot; schedule learning; dynamic
environment; event prediction; multisensory data

I. INTRODUCTION

One of the primary challenges in service robotics is to
allow robots to adapt to open-ended dynamic environments
where they need to interact with naïve users with a little
understanding of the strengths and limitations of the robotic
systems [1]. To operate safely and efficiently in these complex
real situations such as private homes, the robot needs to
understand the environments and ongoing events.

To illustrate some of these problems, we employ the
following example of a schedule that could happen in the
morning at home, there existing a child going to an elementary
school. A robot has to perform a role of a mother who takes
care of the schoolchild. It can wake up the child, guides
him/her to do some exercise and to take a shower, gives a
breakfast, helps to wear appropriate clothes, and informs to
pack study supplies or textbooks.

One solution to this problem is to hand-code the robot
behaviors following a predefined user’s schedule and home
environment dataset. But there is no guarantee that the child
would finish each task for exact time or every object would be
located at the right position. Therefore, it is not feasible to
hand-code all possible behaviors and to predict all possible
exceptions under these complex real situations that the robot is
faced with.

We aim to solve the problem by developing an autonomous
schedule learning platform dubbed as Schedulebot. The bot
freely moves around home and perceive the spatial information
and objects, or follows family members and observe their
activities and conversation. Based on the spatiotemporal data,
the robot understands current events and perform right actions.

II. EXPERIMENTAL SETUP

To show how mother and child normally spend a morning
time to a robot, we built a pseudo-real home environment that
family members and a robot can interact with each other based
on the morning at home scenario as described in Table 1. The
home space is divided into a bedroom, living room, kitchen,
bathroom, and entrance, as shown in figure 1.

The robot can move all of the home spaces by wheels and
observe the environments and people using a wide range of
sensors, including RGB and depth camera, infrared and laser
sensors, odometer, and microphones.

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
<th>Location</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>7:00</td>
<td>Wake up</td>
<td>Bedroom</td>
<td>Move, Speech, …</td>
</tr>
<tr>
<td>7:05</td>
<td>Exercise</td>
<td>Living room</td>
<td>Recognize, Speech, …</td>
</tr>
<tr>
<td>7:13</td>
<td>Washing</td>
<td>Bathroom</td>
<td>Move, Speech, …</td>
</tr>
<tr>
<td>7:20</td>
<td>Breakfast</td>
<td>Kitchen</td>
<td>Recognize, Speech, …</td>
</tr>
</tbody>
</table>
III. ONGOING RESULTS

A. Understanding Visual Scene

The robot needs to understand what it observes from multimodal sensors. To integrate various types of sensory information, word captioning for each data will be used.

First of all, we generated captions that describe visual scenes recorded by an RGB camera. To understand visual scenes from real situations, the localization (e.g. object detection) and description task needs to be jointly performed. We used a fully convolutional localization network architecture [2] that processes an image with a single, efficient forward pass, requires no external regions proposals, and can be trained end-to-end with a single round of optimization. The architecture is composed of a convolutional network, a novel dense localization layer, and recurrent neural network language model that generates the label sequences.

Figure 2 shows the generated captions for example visual scene. There are several bounding boxes that are detected objects and people, and captions describe each box are generated. A robot then can integrate various information from the captions and better understand the visual scene.

![generated captions](image)

**Fig. 2.** Generated captions that describe the environment

B. Event Prediction

The robot also needs to predict the current and next events to do appropriate tasks. This event prediction generally will be processed based on the environmental understanding. However, in advance, we tried to perform this task by using a pseudo schedule data vectors, which is a molecular evolutionary architecture for cognitive learning and memory.

Predicted events, activities, and robot actions following the results are shown in figure 3. Every event in the scenario is scheduled reasonably except one event: studying event is omitted because the number of data is very small since it happens rarely.

![event prediction](image)

**Fig. 3.** Event prediction results and robot actions

ACKNOWLEDGMENT

This work was partly supported by the Institute for Information & Communications Technology Promotion (RO126-16-1072-SW.StarLab), Korea Evaluation Institute of Industrial Technology (10044009-HRI.MESSI, 10060086-RISF), and Agency for Defense Development (UD130070ID-BMKR) grant funded by the Korea government (MSIP, DAPA).

REFERENCES


Evaluation of User Interface Design using Psychophysiological Measures

An EEG Study on the Formation of User’s First Impression on an Interface Design
With a Focus on Aesthetics and Expected Usability

Bonseung Koo
Graduate Program of Cognitive Science
Yonsei University
Seoul, Korea

Nahye Kim
Graduate School of Information
Yonsei University
Seoul, Korea

Kwangsu Cho
Graduate School of Information; Graduate Program of Cognitive Science
Yonsei University
Seoul, Korea

Abstract—Although traditional UX evaluation methods have focused on the experience users undergo when they actually use a service or product, the first impression that users obtain before actually using it is also an important factor to take into consideration. Previous studies have found that users form the first impression about an interface design in just a few seconds and quickly decide on whether to stay on or leave the interface. Furthermore, the first impression that is formed at the moment that users face the interface influences subsequent decisions they make, having impact on their task performance and preferences later on. Regarding such a significant role of the first impression, the current research suggests a brain based measurement method, EEG, as a solution to measure the first impression of users. Specifically, with EEG as an experimental method, this study aims at judging user’s first impression on an interface and understanding the underlying neural mechanisms for the process by which the first impressions are formed.

Keywords—EEG; ERP; First impression; Interface design; UX; Aesthetics; Expected Usability

I. INTRODUCTION

Common UX evaluation methods can be largely divided into two dimensions: behavioral and attitudinal methods [1]. First, behavioral methods refer to those that measure what people ‘do’ with a certain product or service, focusing on people’s behaviors. Examples of behavioral methods are eye tracking, clickstream analysis, and A/B testing. Second, on the other extreme, there are attitudinal methods, which are what people ‘say.’ With attitudinal methods, researchers intend to understand people’s stated beliefs by mainly collecting self-reported information. Attitudinal methods include card sorting, surveys, and focus group interview.

Each of the traditional UX assessment methods mentioned above has disadvantages. First, attitudinal methods have a high risk of social desirability bias, which refers to a tendency of respondents to answer questions or express their opinions in a way that can be viewed as favorable by other members of the society [2]. This can cause serious problems in UX evaluation because without participants’ honest opinions about products to be launched, researchers are not able to accurately figure out the strong points to emphasis on and weak points to improve in their products. The second category, behavioral methods, reflects only one aspect of complex human behavior and cannot provide much insight on emotional states. In case of eye tracking, it only records information about fovea vision, which is only 2% of the entire human visual field [3]. Moreover, although behavioral methods can measure objective, accurate data on how humans behave with a certain product or in a certain situation, the reasons for that behavior or user’s feelings while behaving in a certain way have to be inferred by the researchers.

To overcome the shortcomings of traditional UX evaluation methods, this study suggests a brain-based measurement method, Electroencephalogram (EEG), as an alternative solution. First, by using EEG, the problems associated with attitudinal methods, which are being subjective responses and inaccurate, recalled memory, can be resolved. EEG is an objective measurement tool because people cannot control their own brain responses, therefore rules out the social desirability bias. Also, EEG measures people’s brain activity while interacting with a product in real time. EEG provides a solution to the limitations of behavioral studies.

II. METHOD

Seventeen healthy right-handed Korean participants (5 females, mean age: 27.47 years, age range: 21-32 years) participated in this experiment. Participants were rewarded 15,000 Won for compensation. All the participants had normal or corrected-to-normal vision and no one had a history of neurological or psychiatric disorders. It was also ensured that no one had any use of medication known to affect the central nervous system.

This study investigated participants’ neural responses to website screenshots. The stimulus set consisted of 4 different versions of the same website. In order to control for the context of website, a web page for a university laboratory was adopted.
Two variables of the website, usability and aesthetics, were manipulated, thereby constituting a 2 (usability: high vs. low) by 2 (aesthetics: high vs. low) structure. The four conditions of the stimuli were: 1) High Aesthetics – High Usability, 2) High Aesthetics – Low Usability, 3) Low Aesthetics – High Usability, 4) Low Aesthetics – Low Usability.

Subjects were seated in front of a computer screen under a dim lighting. Participants were about 90cm away from the computer screen. Prior to starting the experiment, experimenters explained to participants about the meanings of each of the question keywords, ‘Aesthetics’ and ‘Usability’, since they would appear on the screen as a single word instead of a question form. Participants were also informed that they would be required to rate aesthetics and usability level of different interface designs presented on the monitor. Therefore, they were instructed to focus on the aesthetics and usability aspects of the design during the experiment. Stimulus presentation was created by E-Prime Software version 2.0.

The electroencephalogram (EEG) was recorded with a 32-channel EGI system (Net Amps 300 with 32 channel HydroCel™ Geodesic Sensor Net, Electrical Geodesics, Inc., Eugene, OR). Electrical activity was amplified with a band-pass filter of 0.1-30Hz at a sampling rate of 1000Hz. The Cz electrode was used as a reference. Data were recorded using EGI’s acquisition software (Net Station version 4.2.4). It was ensured that impedances were kept under 50 kΩ and they were examined for electrode bridging.

### III. RESULTS

#### A. N100 (80-140ms Epoch)

For N100 window, there was a main effect of aesthetics at the frontal (F3, Fz, F4), central (C3, Cz, C4), and parietal (P3, Pz, P4) regions [frontal: F(1, 16) = 9.71, p < .05; high aesthetics: -1.25V; low aesthetics: -0.022V, central: F(1, 16) = 8.56, p < .05; high aesthetics: -1.39V; low aesthetics: -0.30V, parietal: F(1, 16) = 11.6, p < .05; high aesthetics: 0.14V; low aesthetics: 1.28V]. There was a larger negativity for high aesthetics compared to low aesthetics interfaces. There was neither significant main effect of usability nor any interaction effect between aesthetics and usability.

#### B. P200 (180-230ms Epoch)

In the time window from 180ms to 230ms encompassing a P200 component, there was a main effect of aesthetics at left frontal (F3) and central (C3, CZ, C4) sites [left frontal: F(1, 16) = 8.27, p < .05; high aesthetics: 2.79V; low aesthetics: 1.11V, central: F(1, 16) = 9.94, p < .05; high aesthetics: 3.31V; low aesthetics: 0.57V]. Interfaces that were high in aesthetics yielded larger positive amplitudes than those low in aesthetics. There was no significant main effect of usability or any interaction effect between aesthetics and usability.

#### C. N300 (260-330ms Epoch)

N300 was modulated by the main effect of usability at frontal (Fz) and central (C3, CZ, C4) sites [frontal: F(1, 16) = 5.37, p < .05; high usability: 2.82V, low usability: 1.78V, central: F(1, 16) = 6.40, p < .05; high usability: 1.04V, low usability: 0.28V]. Interfaces that are low in usability yielded more negative amplitudes compared to those with high usability. There was neither significant main effect of aesthetics nor any interaction effect between aesthetics and usability.

### IV. CONCLUSION

In this study, an ERP experiment was conducted in order to investigate momentary brain responses that capture users’ first impressions on user interfaces, with a specific focus on aesthetic and expected usability. The results of this study imply that when evaluating an interface design, users’ judgment about usability and aesthetics based on the first impression of the design is completed within 500ms time period. The time course of this process starts from evaluation of aesthetics until around 200ms, and the evaluation of expected usability emerges at about 300ms. The results of the current study do not show any interaction effect between usability and aesthetics. It is difficult to conclude on which one of aesthetics or usability is more important from this result. However, it could be argued that since aesthetics is the first component that is assessed at the earliest time period, leaving a good impression on how beautiful the interface appears might play an important role in making an immediate positive attitude.

### REFERENCES

Cambot: A Visual Conversation Robot for Interactive Engagement

Kibeom Kim1,3, Jin-Hwa Kim2,3, Byoung-Tak Zhang1,2,8
Surromind Robotics1
Cognitive Science Program2
Cognitive Robotics and Artificial Intelligence Center3
Seoul National University
Seoul, Republic of Korea
Email: {kbkim,jhkim,btzhang}@bi.snu.ac.kr

Abstract—To achieve human-level artificial intelligence, it is crucial to develop algorithms which handle human-like visual and linguistic information. One of promising solutions is to use Multimodal Residual Networks (MRN) for the multimodal residual learning in assumption of visual question-answering tasks. It extends the idea of the deep residual learning, which learns joint representation from vision and language information effectively. While the MRN is handling with multidisciplinary problems of vision, language and integrated reasoning, a visual conversation robot can be a bridge to interact with humans. Cambot can be instantiated in any platform including robots, desktops and tablet PCs, which have a camera and microphone, engaging natural environmental situations of visual conversation for human interactions.

I. INTRODUCTION

We have been seeking human level artificial intelligence (AI) technology. For the technology, the studies in AI have been researched narrow subject like image recognition [1], [2], [3] and language model [4], [5], [6]. But, we need to make a robot which can handle multidisciplinary problems of vision, language and integrated reasoning like a person. Like the cambot, RI-MAN [7] and Sony’s robot pet, AIBO [8], are interactive robots. But these don’t have visual and linguistic ability. The robot that use a core technology as Multimodal Residual Networks (MRN) [9] that is great performance at visual question-answering tasks [10] can understand vision and linguistic information. One of promising solutions is to make a robot which can handle multidisciplinary problems of vision, language and integrated reasoning, a visual conversation robot can be a bridge to interact with humans. Like the cambot, RI-MAN [7] and Sony’s robot pet, AIBO [8], are interactive robots. But these don’t have visual and linguistic ability. The robot that use a core technology as Multimodal Residual Networks (MRN) [9] that is great performance at visual question-answering tasks [10] can understand vision and linguistic information. One of promising solutions is to make a robot which can handle multidisciplinary problems of vision, language and integrated reasoning, a visual conversation robot can be a bridge to interact with humans. Like the cambot, RI-MAN [7] and Sony’s robot pet, AIBO [8], are interactive robots. But these don’t have visual and linguistic ability. The robot that use a core technology as Multimodal Residual Networks (MRN) [9] that is great performance at visual question-answering tasks [10] can understand vision and linguistic information. One of promising solutions is to make a robot which can handle multidisciplinary problems of vision, language and integrated reasoning, a visual conversation robot can be a bridge to interact with humans.

II. MULTIMODAL RESIDUAL NETWORKS

MRN consists of multiple learning blocks, which are stacked for deep residual learning. Denoting an optimal mapping by \(\hat{H}(q,v)\), we approximate it using

\[
H_1(q,v) = W_{q'} q + F^{(1)}(q,v).
\]  

The first (linear) approximation term is \(W_{q'} q\) and the first joint residual function is given by \(F^{(1)}(q,v)\). The linear mapping \(W_{q'}\) is used for matching a feature dimension. It is defined the joint residual function as

\[
F^{(k)}(q,v) = \sigma(W_2^k q) \odot \sigma(W_1^k v) + W_{q,v} q \odot v
\]

where \(\sigma\) is tanh, and \(\odot\) is element-wise multiplication. The question vector and the visual feature vector directly contribute to the joint representation.

For a deeper residual learning, we replace \(q\) with \(H_{l-1}(q,v)\) in the next layer. In more general terms, (1) and (2) can be rewritten as

\[
H_L(q,v) = W_q q + \sum_{l=1}^{L} W_{q,v}^{l-1} F^{(l)}(H_{l-1}, v)
\]

where \(L\) is the number of learning blocks, \(H_0 = q, W_q = \Pi_{l=1}^{L} W_q^{l}, \) and \(W_{q,v} = \Pi_{m=1}^{L} W_{q,v}^{m}\). Notice that the shortcuts for a visual part are identity mappings to transfer the input visual feature vector to each layer (dashed line). At the end of each block, it is denoted \(H_l\) as the output of the \(l\)-th learning block, and \(\odot\) is element-wise addition.

III. CAMBOT

Here, we suggest a cambot as a platform for visual conversation. It consists of three parts, such as a robot hardware, a web server and a visual question-answering (VQA) server. It is illustrated on fig.2. The robot hardware is equipped with

![Cambot: A Visual Conversation Robot for Interactive Engagement](image-url)
microphones, cameras and speakers like a human. From the microphones and cameras, the robot gets audio signals and images from the human. If a person ask a question to the robot, the Cambot take a picture and the audio signals, and the signal is converted to Korean text. We used the Google Speech Recognition APIs for converting the signal to the text. Then, the data sends to web server that is second part. If the Cambot will speak when returned data is received from web server. Fig. 3, is a scene depicting a conversation that a man ask a question and the Cambot has an answer.

The web server which is made for working on any platforms like desktops and tablet PCs connects the robot hardware and the VQA server. In this part, a sentence of input data is translated Korean into English via the Japanese using the MS Bing translator APIs, as the VQA server is learned English. And the data from the third part is translated English into Korean. If the input sentence translates Korean into English directly, the translation is not clean. But Japanese and Korean is better to translate because the sentence structure is similar. Furthermore, translation of Japanese and English are cleaner than Korean and Japanese because of the large amount of data.

The last part is the VQA server. The server which is using the MRN is received the input data that is an English sentence and image. After process, output data will return to web server as an answer. The MRN is learned the VQA dataset [10] which are collected via Amazon Mechanical Turk from human subjects, who satisfy the experimental requirement. It includes 614,163 questions and 7,984,119 answers, since ten human subjects, who satisfy the experimental requirement. It is an advanced technology that handles a sentence and image. And the web server which connects a robot and the VQA server translate Korean into English and English into Korean. Using the cambot, it is possible to play with children and teach a foreign language to them. Furthermore, we expect this technology may help blind people to make it easy to find something by talking with them. Although the cambot handle captured image, the cambot will process video data and other sensing like a human.

IV. CONCLUSION

The Cambot with MRN is state-of-the-art in visual question-answering. It is an advanced technology that handles a sentence and image. And the web server which connects a robot and the VQA server translate Korean into English and English into Korean. Using the cambot, it is possible to play with children and teach a foreign language to them. Furthermore, we expect this technology may help blind people to make it easy to find something by talking with them. Although the cambot handle captured image, the cambot will process video data and other sensing like a human.

REFERENCES

Glassbot: Personalized Wearable Agents
Learning from Everyday Human Behaviors

Sang-Woo Lee∗, Chung-Yeon Lee∗, Dong-Hyun Kwak† and Byoung-Tak Zhang∗†
∗School of Computer Science and Engineering
†Interdisciplinary Program in Neuroscience, Seoul National University
Seoul National University
Seoul, 08826, South Korea
E-mail: \{slee, cylee, dhkwak, btzhang\}@bi.snu.ac.kr

Abstract—This paper introduces Glassbot, the agent on glass-type wearable devices with camera and audio sensors. We want to train Glassbot continuously in a wearable device by rapidly adapting deep neural networks from sensor data streams of user behaviors. In this paper, we describe our early works on dataset and online learning algorithms for Glassbot. We also discuss the future work of predictive learning and scheduling on Glassbot.

I. INTRODUCTION

It is essential for building human-aware intelligent agent that can learn from human behaviors in the real world. Glass-type wearable devices, including Google Glass and Narrative Clip, have merits on this problem. It is noticeable that these devices can see and hear what the device user sees and hears; this property differentiates them from the classical agents in personal computers or the smartphones.

We introduce the concept of Glassbot, the agent on Glass-type wearable devices. We believe Glassbot is promising for the next-generation digital assistant, because video stream collected through wearable devices have a wealth of information, and recent deep learning technique can handle this to make the agent smarter. To implement Glassbot, we are paying attention to three following technique; 1) Acquisition of knowledge from real stream data, 2) continuous learning of knowledge, 3) predictive learning over daily life. Potential applications of Glassbot is as follows; 1) Context recognition, 2) Scheduling, 3) question answering on the context, 4) automatic diary writing.

We are currently doing one early work of Glassbot. In this paper, we describe our early works on dataset and online learning algorithms for Glassbot. We also discuss on the future work of predictive learning and scheduling on Glassbot.

II. LIFELOG DATASET

The environment Glassbot encounters has two properties. First, there are hidden high-level contexts in the raw sensory stream, for example, an first-person view video stream recorded during a dating includes various types of high-level contexts, although the data is only a stream of pixels and audio signals. Second, the data streams are often non-stationary, for example, the life patterns of the weekday and weekend are different.

We collected Google Glass lifelog dataset recorded over 46 days from three participants [1] (Figure 1). The 660,000 seconds of the first-person view video stream data reflects the behaviors of participant including the indoor activities, such as ‘studying in the library’ or ‘watching TV in the house’, and the outdoor activities, such as ‘walking on the road’ or ‘waiting for the arrival of the bus’. The participants were asked to notate what they were doing and where they were in real-time by using a life-logging application installed on their mobile phones. In this study, location, sub-location, and daily activity

Fig. 1. Lifelog dataset collected through Google Glass

<table>
<thead>
<tr>
<th>Location</th>
<th>Sub-location</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>university (147034)</td>
<td>classroom (101844)</td>
<td>studying (90330)</td>
</tr>
<tr>
<td>outside (130754)</td>
<td>home-room (86588)</td>
<td>watching (35387)</td>
</tr>
<tr>
<td>home (97180)</td>
<td>subway (35284)</td>
<td>eating (60725)</td>
</tr>
<tr>
<td>restaurant (22190)</td>
<td>bus (34120)</td>
<td>working (204131)</td>
</tr>
</tbody>
</table>

TABLE I

<table>
<thead>
<tr>
<th>Instances (sec/day)</th>
<th>Number of class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>Test</td>
</tr>
<tr>
<td>A 103420 (13)</td>
<td>17055 (5)</td>
</tr>
<tr>
<td>B 242845 (10)</td>
<td>91316 (4)</td>
</tr>
<tr>
<td>C 144162 (10)</td>
<td>61029 (4)</td>
</tr>
</tbody>
</table>

TABLE II

<table>
<thead>
<tr>
<th>Location</th>
<th>Sub-location</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>office (196839)</td>
<td>office-room (182884)</td>
<td>working (204131)</td>
</tr>
<tr>
<td>university (147034)</td>
<td>classroom (101844)</td>
<td>studying (90330)</td>
</tr>
<tr>
<td>outside (130754)</td>
<td>home-room (86588)</td>
<td>watching (35387)</td>
</tr>
<tr>
<td>home (97180)</td>
<td>subway (35284)</td>
<td>eating (60725)</td>
</tr>
<tr>
<td>restaurant (22190)</td>
<td>bus (34120)</td>
<td>working (204131)</td>
</tr>
</tbody>
</table>
Fig. 2. Dual memory architecture (DMA) for lifelong learning of Glassbot agent

are labeled. A frame image of each second is used as one instance. Table I summarizes the dataset statistics and Table II presents the distribution of the five major classes in each class type.

III. LIFELONG LEARNING

We are interested in adapting the context-aware activity recognizer continually and rapidly from human behaviors gathered through wearable devices. To treat these properties, two algorithmic techniques are required. First, the deep learning method is necessary to handle raw-level data efficiently [2]. Second, an online learning algorithm is required to keep track of fast-changing life patterns of user behavior [3].

To address these issues together, we utilized the concept of complementary learning systems (CLS) theory a framework that suggests a dual learning system structure in the brain [4]. According to the CLS theory, there are two critical areas in the brain that affect online learning: the neocortex and hippocampus, which complement each other’s functionality.

Inspired by the CLS theory, we propose a dual memory architecture (DMA) (Figure 2) [1]. The DMA trains two memory structures: one is an ensemble of DNNs, and the other consists of a shallow network that uses hidden representations of the DNNs as input. These two memory structures are designed to use different strategies. The ensemble of DNNs learns new information to adapt its representation to new data, whereas the shallow network aims to manage non-stationary distribution and unseen classes more rapidly.

In our experiments, the proposed DMA outperformed other online learning methods on two datasets: the CIFAR-10 image-stream dataset and the lifelong dataset. Comparative learning methods often failed to keep the information of old data. Figure 3 illustrate the story of daily life in the lifelog dataset and the effectiveness of DMA on this scenario. As DMA can learn deep representation of new events, DMA outperforms comparative results and robustly learns from non-stationary environment.

IV. PREDICTIVE LEARNING AND SCHEDULING

We want to learn the daily story of the user and predict future daily activity based on the knowledge of the daily story. The typical application of predictive learning is scheduling. By understanding the current context and extracting a series of daily schedule from lifelog dataset, digital assistant can recommend and predict user behaviors [5]. Predicted user behaviors can be used for question answering about user’s schedule or multimodal context [6].

Currently, we collect more data for simulating the scheduling service of Glassbot. In addition to video data, the schedule of the user will be augmented. The continuity of the dataset will also be enhanced.

ACKNOWLEDGMENT

This work was supported by the Naver Corp. and partly by the Korea Government (IITP-R0126-16-1072-SW.StarLab, KEIT-10060086-HRI.MESSI, KEIT-10044009-RISF, BMRR-UD130070ID).

REFERENCES

Jibobot: A Personal Assistant Robot with Social Motions

Eun-Sol Kim, Jiseob Kim, Donghyun Kwak, Kyoungh-Woon On and Byoung-Tak Zhang
Department Computer Science and Engineering
Seoul, 08826, South Korea
Email: {eskim, jkim, dhkwak, kwon, btzhang}@bi.snu.ac.kr

Abstract—For a personal assistant robot to be attractive, it is crucial that it generates social behaviors that are shared with human users. However, the robot body is different from the human body and mapping human behavior into robot behavior is a challenge. Here we develop a reinforcement learning method for personal robots to generate human-adaptive social behaviors from continuous feedback. We first build a basic motion generation model using a Gaussian Process Dynamic Model (GPDM) that is trained on a human activity dataset. Then, we continuously personalize the generated motions with reward feedback through a Bayesian optimization method for reward regression. The method is evaluated using a personal robot platform to generate seven kinds of robot motions using real-valued rewards, the system finds a function which estimates rewards about the motion from a human. Using these real-valued rewards, the system gets continuous feedback and maps the human behavior into robot behavior. To do this, the system gets a reward value from a human and then shows a new robot motion to be evaluated. After a small number of iterations, getting a reward from a human and showing a new robot motion, the system finds a reward regression function to find the maximum reward and the corresponding motion.

A. Basic Motion Generation

We suggest GPDM [2], [3] as a method to generate basic robot motions of each status class. The reasons GPDM is used are 1) it compresses nonlinear time series data into a low-dimensional latent space, smoothly; 2) it does not require a large number of datasets for training; 3) the dimension size of the latent space can be chosen without changing the algorithm.

A GPDM is composed of a low-dimensional latent space with associated dynamics, and a map from the latent space to an observation space. The model parameters, related to the dynamics and the mapping, are marginalized out in closed-form using Gaussian Process priors [1]. Therefore, smooth low-dimensional representations of the nonlinear time series data can be obtained with GPDM.

In our problem setting, the GPDM is used to find basic robot motions (low-dimensional latent space) trained on the human activity dataset (nonlinear time series data).

B. Human-Adaptive Robot Motions

After finding basic robot motions modeled after human activities by GPDM, the system keeps finding better robot motions which can get positive responses from a human. In this paper, we suggest a Bayesian optimization method as an exploration method to find better robot motions.

The main idea of finding an optimal robot motion is getting rewards about the current motion from a human and changing the robot motion until the optimal robot motion is discovered. The system gets real-valued rewards from a human about
Fig. 1. The overall framework for generating human-adaptive robot motions. (a) Human activity data is collected using Kinect and ROS. Each human motion sequence consists of 45 absolute coordinates of 15 joints and contains 300-400 frames. (b) The human activity data is compressed into low-dimensional latent space by GPDM. Here, we assume the robot platform has three degrees-of-freedom, so the latent space is three-dimensional. In this step, a three-dimensional sequence is generated and is used as the basic motion of the robot. (c)-(d) Human evaluates the motion in terms of a real-value. The system gets the rewards over and over and finds a mapping function from the robot motions to the human rewards. After estimating the reward functions, the system finds the maximum value of the reward and the correspond robot motion which is most attractive to the human-user.

C. Experimental Results and Conclusion

In Figure 2, a sequence of inspection points and corresponding robot motions are described through Bayesian optimization. As the procedure proceeds, we can see that various motions are explored such as changing motors or velocities.

In this work, we showed a new reinforcement learning method to generate human-adaptive robot motions. The generated robot motions are learned from continuous human feedback, therefore the motions can be regarded as the robot motions having the maximum human rewards. The suggested method could have novelty as the robot motions are evolving in a direction getting more positive human rewards automatically. Also, this method easily applies to other robot platforms which have different robotic structures.

REFERENCES


Pandabot: Multimodal Story Learning with Dynamic Memory Construction

Yu-Jung Heo, Eun-Sol Kim, Kyoung-Woon On and Byoung-Tak Zhang
Department of Computer Science and Engineering
Seoul National University
Seoul 08826, Korea
{yjheo, eskim, kwon, btzhang}@bi.snu.ac.kr

Abstract—We consider a challenging problem, video question and answering, based on multimodal sequential information. We suggest a memory-based machine learning algorithm which temporally combines three kinds of information, such as image stream, subscriptions and audio signals, of videos and infers an answer for a given question based on the constructed memory.

Compared to conventional methods for question and answering, the suggested method can construct dynamic memory representations of multimodal sequential data. Also it can generate subjective answers rather than one-word objective answers for the questions of the human. For the experiments, 4000 pairs of question and answering data are collected by Amazon Mechanical Turk for five fairy tales of the child educational animation video. Experimental results on the collected question and answering dataset demonstrate meaningful improvements of the accuracy of the answers.

Video question and answering is a challenging problem, because it is needed to infer high-level semantics based on the context of the video. In this paper, we propose a memory-based story learning method which can deal with multimodal sequential information about the story of video. Using the constructed multimodal dynamic memory, video question and answering problem is resolved.

The suggested method extends Dynamic Memory Network(DMN)[1] which focuses on language-based question and answering problems. DMN constructs episodic memory just using text information such as description and question sequence and generates relevant answers given a memory. Even though the model shows significant performance on question and answering, it is hard to apply to the video. Because, in video question and answering problem, not only text information explaining the situation but also visual and auditory information is crucial to understand overall context in given situation.

Therefore, in this paper, we combine three kinds of modalities, such as image stream, subscriptions, and audio signals, of the video to construct episodic memory. Also, as constructing each episode, we consider each question time stamp when a user asked each question watching the video. We focus on a point in time for question and answering, so that make the episodic memory dynamic and compacted.

A. Input Module

The input module combines three modalities, such as image stream, subscription and audio signals, of given video. Raw data for each modality is properly preprocessed to have high-level representations. We use a gated recurrent network(ReLU)[2][3] to encode sequential information of preprocessed inputs $I^t$ with time $t$. The equations of mechanism how to GRU network works are as follows.

$$ z_t = \sigma(W(z)(I^t) + U(z)h_{t-1} + b(z)) $$

$$ r_t = \sigma(W(r)(I^t) + U(r)h_{t-1} + b(r)) $$

$$ \tilde{h}_t = \tanh(W(I^t) + r_t u h_{t-1} + b(h)) $$

$$ h_t = z_t c(I_{t-1}^e) + (1 - z_t) \sigma(h_t) = GRU(I^t, h_{t-1}) $$

B. Question Module

The question consists of a sequence of words. The question module encodes these words using the recurrent neural network same as input module. Hidden state of $h$th word of question is given by $q^e = GRU(\text{embed}[w^q_{e}], q^{e-1})$. Final hidden representation $q^{R(e)}$ is applied to construct episodic memory module $m_t$.

C. Episodic Memory Module

The episodic memory is constructed by iteration step which updates previous memory states over the input and question. It can be considered to apply change of attention on question and answering. The update equation of the episodic memory is $m_t = GRU(e^t, m_{t-1})$

D. Answer Module

The answer module generates hidden representations of answers $a = [a^1, a^2, ..., a^N]$, in which $N$ is the number of answers. Each element of $a$ means the final hidden representation of each answer, encoded by GRU module. $t$th hidden state is given by $a^t = GRU(\text{embed}[w^q_{a}], a^{t-1})$. With respect to final episodic memory module $m_{T(M)}$, the model output best answer $A^t = \min ||(m_{T(M)}^t - a^t)||^2$.

E. Data

As experimental dataset, we used educational animations made by Pinkfong, No.1 grossing education app in 109 countries. We first selected five fairy tales, such as Hansel and Gretel, Snow White and the Seven Dwarves, The Little Mermaid, The Wolf and the Seven Sheep, and The Three Little Pigs. Then, we collected 800 question and answering pairs via
Fig. 1. An overall structure of Memory Network with dynamic memory construction for multimodal story learning

Amazon Mechanical Turk, a crowd-sourcing platform. Each QA pair has time stamp, which represents the timing of the contents related to the question. The question and answering pairs can be divided three types: first, question and answering about story (e.g., Why did Snow White fall? She ate the poison apple.), second, question and answering about visual information (e.g., What color is the poison apple? The apple is red.), third, question and answering about emotional feeling (e.g., What feeling does the Queen have? She is jealous.).

F. Preliminary results

We conducted preliminary experiments to verify suggested dynamic memory construction make episodic memory compacted. As the experiment setting, The text information of input and question module are embedded by pre-trained word embedding representation using GloVe[4]. We divided the collected dataset into two sets: one half for training set and the other for test set. Table I shows some examples of the test results.

G. Conclusion and Future works

In this paper, we proposed a memory-based story learning method for video question and answering problems. Also, we conducted preliminary experiments using dynamic memory construction and demonstrate the preliminary results. As the following experiments, we will show the results from multi-modal sequential data of the video.

<table>
<thead>
<tr>
<th>TABLE I</th>
<th>SOME EXAMPLES OF THE TEST RESULTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>TYPE</td>
<td>SENTENCE</td>
</tr>
<tr>
<td>Question</td>
<td>Does the hunter kill Snow White?</td>
</tr>
<tr>
<td>Answer</td>
<td>No he let her go into the Forest.</td>
</tr>
<tr>
<td>Prediction(Ours)</td>
<td>no hunter doesn't kill snow white.</td>
</tr>
</tbody>
</table>

| Question| Why is the queen angry?        |
| Answer  | Because the mirror tells her Snow White is the most beautiful. |
| Prediction(Ours)| the mirror replied snow white is the most beautiful.|

REFERENCES

Abstract—It is well known that two complementary learning modules are important to achieve human-level language understanding: one gradually acquires structured representations of knowledge from language, the other quickly learns the episodic memory composed of individual knowledge. In this paper, we suggest a new machine learning model which combines symbolic and neuronal approaches to construct the complementary learning model. The suggested model extracts symbolic knowledge from natural language sentences, then the symbolic knowledge is embedded into a real-valued continuous vector space, a neural representation. The neural representation implies the meaning and the correlations between the symbolic knowledge, also generalize the patterns among the knowledge. At the same time, the model rapidly learns to predict a specific knowledge which is supposed to be based on the reasoning. As an application of the suggested method, we conduct a challenging problem, a question and answering, which needs to understand the context of the language input and reason answer for the given question based on the context.

Complementary learning system (CLS) is a great theoretical base to understand the mechanism of human learning and memory. According to the theory, in neocortex, semantic knowledge is gradually constructed using information from episodic memory. At the same time, in hippocampus, the episodic memory is rapidly constructed using semantic knowledge structure[7], [8]. [9]. Getting intuitions from the CLS theory, we consider a model combining a symbolic approach (from Artificial intelligence) and a neural representation approach (from the recent machine learning) to build complementary system for learning human language.

In this paper, we suggest a new model to construct the neuroknowledge based complementary learning architecture to understand and reason about knowledge written in human language. For demonstration, we will show a novel challenging problem, question and answering task which needs reasoning answer from the text input and question.

A. Model description

The suggested model consists of four parts: a symbolic knowledge extraction module, a neuroknowledge representation module, an episodic memory module, an answer module. In the symbolic knowledge extraction module, the symbolic knowledge triplet, <subject, relation, object>, is automatically extracted from the input text. Then the neuroknowledge representation module learns generalized neural representation of each knowledge triplet. The episodic memory module learns to predict to pick specific knowledge from a trigger, question for example, then finally output will be comes out from the answer module. A high-level illustration of the model is shown in Figure ??.

1) Symbolic knowledge extraction module: To extract symbolic knowledge automatically from the text, we use open information extraction (OpenIE) which can identify entities (subject and object) and relations from natural sentences [7]. For example, given the sentence, “McCain, fought hard against Obama, but finally lost the election,” an OpenIE system may extract two triplets, <McCain, fought against Obama>, and <McCain, lost, the election>. Using this technique, multiple symbolic knowledge triplet are obtained from natural text input.

2) Neuroknowledge representation module: From the symbolic knowledge triplet, the neuroknowledge representation module learns generalized neural representations of each symbolic knowledge triplet. There are several approaches embedding knowledge triplets to neural representation, but most of them more focused on entity embedding which toward to reflected relation in a fixed number of relation environment [7], [9]. In the suggested model, we use factored high-order Boltzmann machine to learn neural representation of knowledge triplets. The factored high-order Boltzmann machine is shown to have nature of capturing correlational structure among inputs [7], [9]. Using this property, we feed Word2Vec [9] representation of <subject, relation, object> triplet as an input, and use hidden representation as neuroknowledge representation. Specifically, to obtain neuroknowledge representation, we firstly embed each component of a triplet, to continuous vector space using word2vec as follows.

\[ e_s = \text{word2vec}(\text{subject}) \]
\[ e_r = \text{word2vec}(\text{relation}) \]
\[ e_o = \text{word2vec}(\text{object}) \]
Fig. 1. Overview of suggested framework. The blue arrow indicates flow of inference and the red arrow indicates back-propagation learning. Briefly, the suggested model extract semantic knowledge triplet from input text, then learn neuroknowledge representation of each triplet. In the episodic memory, all knowledge is stored in a form of neuroknowledge representation and key for predicting appropriate knowledge to be learned. Using this key, specific knowledge for question is extracted and based on that knowledge, the answer to be predicted.

The word2vec() is the word2vec function. After that, the neuroknowledge representation $h_n$ of specific $<e_s, e_r, e_o>$ triplet is determined by

$$h_n = \sigma \left( W_{af} \cdot (W_{rf} \cdot e_s) \circ (W_{rf} \cdot e_r) \circ (W_{rf} \cdot e_o) + \text{bias} \right)$$ (2)

Here, $W_{af}, W_{rf}$ are weight matrices, $h_n, e_s, e_r, e_o$ and bias are vector and $\sigma$ is an activation function. Also, $\cdot$ is the dot product and $\circ$ is the element-wise multiplication.

3) Episodic memory module: In the episodic memory module, the set of symbolic knowledge embedded in neuroknowledge representation is stored sequentially and retrieved based on the trigger, such as questions. The episodic memory module rapidly learns to predict appropriate knowledge to be retrieved corresponding to the question. This process imitate the pattern completion process in hippocampus of brain and is done by two steps. First, a knowledge-key $k_{knowledge}$ is generated by

$$k_{knowledge} = f_{knowledgekey}(e_q)$$ (3)

Here, $e_q$ is an embedding vector of a question and the function $f_{knowledgekey}$ is the multiplerayer perceptron (MLP). This function can be replaced by any other differential functions. Using the generated knowledge-key $k_{knowledge}$, a key-value look-up over the neuroknowledge in episodic memory is performed to retrieve appropriate neuroknowledge $h_n'$,

$$P(h_n'|e_q) = \frac{\exp(k_{knowledgekey}^\top h_n')}{\sum_k \exp(k_{knowledgekey}^\top h_n')}$$ (4)

$$h_n' = \text{argmax} P(h_n'|e_q)$$

4) Answer module: To generate an answer for a question, we first extract $<\text{subject} (e_s), \text{relation} (e_r), \text{object} (e_o)>$ triplet again from the retrieved neuroknowledge $h_n'$. Then the answer $a$ is determined by

$$a = f_{answer}(e_s || e_r || e_o || e_q)$$ (5)

The notation $[a|b|c|d]$ denote the concatenation of vector $a$, $b$, $c$ and $d$. The function $f_{answer}$ used here is the softmax function. Also, $f_{answer}$ can be replaced by other functions.

B. Expected experimental results

To demonstrate the performance of the suggested model, the Q/A experiment will be conducted based on two different dataset: the Facebook bAbI dataset & educational video Q/A dataset. The bAbI dataset is well-known dataset for testing a model’s ability to reason over facts. educational video Q/A dataset is collected questions and answers based on video contents mad by Pinkfong. With these experiments, we will show better results compared with other existing approach.

REFERENCES

Cafebot:
A Conversational Cashier Robot in Korean Cafes

Cheolho Han, Kyoung-Woon On, Eun-Sol Kim, and Byoung-Tak Zhang
Department of Computer Science and Engineering
Seoul National University
South Korea
{chhan, kwon, eskim, btzhang}@bi.snu.ac.kr

Abstract—This paper discusses developments on Cafebot, a conversational cashier robot in Korean cafes. Cafebot has conversations with customers, which requires modeling the intention and response of both speakers, Cafebot and customers. The human voice is not directly understandable by robots, so it must be converted to a text. The text is also not adequate for training robots; we need to change it to digital numbers. Then Cafebot’s conversational model learns the intention and response represented in those numbers. Through these processes, Cafebot gets to talk to customers in cafes. This paper introduces several methods that are available in some primary processes.

Keywords—conversational agents; cognitive robots; bag of words; doc2vec; nearest neighbors; hidden Markov models; support vector machines; hypernetworks

I. INTRODUCTION

Cafebot [1-4] is a conversational cashier robot in Korean cafes. Cafebot has conversations with customers, which requires modeling the intention and response of both speakers, Cafebot and customers. It can be achieved by applying several methods in preprocessing, which converts the speech to the vector, and in modeling, by which the robot learns how to talk. This paper illustrates those methods.

II. PREPROCESSING

A. Hand-crafted Features

The utterance feature (UF) [2, 3] is introduced to discretize a variable which represents each speech of speakers. We made several standards to assort sentences into utterance features. Table 1 shows groups of criteria for classifying utterance features. ‘Menu’, ‘Price’ and ‘Option’ are conditions which categorize sentences whether they include some of menu, information about price or optional choices. The word ‘Action’ signifies activities like paying with the credit card or signing. Furthermore, ‘Complete/Incomplete’ of conditions’ Group 2 reflects one characteristic of Korean language which has structural components of sentence notifying ending of sentences. These standards and information about speakers are used to make utterances into utterance feature which has discrete and finite value.

B. Bag of Words

The Bag-of-words (BoW) model [1] is used for feature extraction of spoken sentences. In the BoW model, a dictionary $T$ is constructed from all of the distinct words in the corpus. Then a sentence is converted to the feature vector where each component is the number of occurrences of each word in $T$.

C. doc2vec

A doc2vec [4] is a method to convert a document to a vector. One is given by genism (genism.models.doc2vec) in Python. The doc2vec separates words by spaces, which leads to differentiating the same words due to different postpositions in Korean. Therefore, each sentence was converted to morphemes before applying the doc2vec.

III. MODEL

Cafebot’s conversational model learns the intention and response represented in vectors. The model has several candidates for itself. Some of them can be applied to both the intention and response, but others only to the intention.

A. Nearest Neighbors

Nearest neighbors (NNs) [4] are one of primitive types of classifiers. NNs can be used to classify the speaker’s intention. The $k$-NNs algorithm determines the class of the input instance, a sentence in conversation, by the majority of the classes of the $k$ previously seen data nearest to the input. The data should be memorized, and they need to be compared to each input, so it may have high memory and time complexity.

B. Hidden Markov Models

Hidden Markov models (HMMs) [2, 3] are used to model sequential data, where some types of data are observable, but...
others are not. HMMs can be used for both the intention classification and the response generation. Training or inference algorithms for HMMs are well known.

C. Support Vector Machines

Support vector machines (SVMs) [1] are a very powerful classifier, and they were actively used in the early 2000s. The most basic form of them is a linear SVM, which maximizes the margin between two classes that are linearly separable. With the kernel trick, SVMs could be applied to broad ranges of data.

D. Hypernetworks

Hypernetworks (HNs) [1] are an extension of graphs. A graph has edges that are a (ordered or unordered) pair of vertices, whereas an HN has edges that are an \( n \)-tuple of vertices. Therefore, a graph is a special case of the HN when \( n = 2 \). HNs are a kind of n-gram models and probabilistic graphical models. They can be used for both the classification and generation.

IV. RESULTS

Through the above processes, we could achieve high accuracy of intention classification and generate responses fairly well. To quantify the generation results, we estimated its accuracy and had people evaluate the naturalness.

A. Intention

The HMM based on hand-crafted features achieved over 90% accuracy in the intention classification (Fig. 1, [3]). The 1-NN based on the doc2vec and the SVM based on the BoW showed 77.4% and 80.14%, respectively.

B. Response

The accuracy of response by treating the generation as the classification was over 50% when the training data were 110 episodes among 130 episodes in total (Fig. 2, [1]). Intention-based HMM showed naturalness score of 4.47 quite close to 6.52 of the real dialogue when the training data were 80% of the whole data (Fig. 3, [3]).

V. CONCLUSIONS

Cafebot has conversations with customers in cafes. It was accomplished by several methods to model the intention and the response.

REFERENCES

Pororobot: Child Tutoring Robot for English Education

Kyung-Min Kim1, Chang-Jun Nan1, Min-Oh Heo1, and Byoung-Tak Zhang1,2
1School of Computer Science and Engineering / 2Cognitive Science and Brain Science Programs
Seoul National University
Seoul, Korea
{kmkim, cjnan, moheo, btzhang} @bi.snu.ac.kr

Abstract—The recent success of machine learning has lead to advancements in robot intelligence and human-robot interaction. It is reported that robots can well understand visual scene information and describe the scenes in language using computer vision and natural language processing methods. Image Question-Answering (QA) systems can be used for human-robot interaction. However, to achieve human-level artificial intelligence based on lifelong learning, a model must deal with real-world environments including dynamic, uncertain, and asynchronous properties based on lifelong learning. In this paper, we propose a prototype system for a video QA robot "Pororobot". The system uses the state-of-the-art machine learning system using dual memory model for implementing robot intelligence. In our scenario, a robot and a child plays a video QA game. Here we demonstrate preliminary results of our system.

Keywords—videoQA; tutoring robot; human-robot interaction; robot intelligence; machine learning;

I. INTRODUCTION

Solving question and answering (QA) problem has been an important theme in artificial intelligence discipline and many computational models have been proposed for few decades. However, in order to achieve human-level artificial intelligence, the models must deal with real-world environments including the dynamic, uncertain, and asynchronous properties based on lifelong learning [1]. Thus, to tackle video QA problem, it is necessary to consider that the concepts in the video dynamically change according to the story. Because of this property, it is hard to answer the questions with only generalized knowledge in a learned model. The model needs to remember not only generalized information but also precise information in short-term context like working memory in human brain.

Here we develop methods i.e. dual deep-learning architectures, for video QA based on two Deep Concept Hierarchies (DCH) [2]. DCH can effectively represent multimodal concepts and efficiently captures the conceptual changes from incrementally observed data. In the architecture, there are two memories that the one learns short-term information with a fixed structure and small memory space and the other learns long-term information with a more flexible structure and large memory space. Thus, they can learn both short-term context and long-term context in the video. As the video stories unfold, each memory independently learns the observed video and have different emergence and evolution phase.

The proposed method is evaluated on a cartoon video series 'Pororo' consisting of approximately 200 episodes with the total playing time of 1232 minutes. We evaluate the model with 100 QA pairs from the video and 20 human evaluators.

II. VIDEO QA SYSTEM OVERVIEW

In our scenario, a child and a robot see the video and they interact each other with the same experience, i.e. watching a video. This could be advantageous for children’s early education [3]. To make this possible considering that the video data are continuous, dynamic, and multimodal, implementing a video QA system requires some issues as follows:

1) Multiple time resolution knowledge
2) Flexible and generalized knowledge representation using multimodal features
3) Continuous knowledge acquisition and update

For solving three issues, we implement a knowledge representation methods based on concepts for enhancing the generalization. Concepts represented with multimodal variable are incrementally learned from continuously increasing data. Especially, as we consider a dynamic environment where children ask questions while observing cartoon videos, we place the two knowledge base in our system.

Figure 1 shows an overview of our video QA system. The QA system can automatically generate the questions from the observed video and answer the questions. Also, because of the
dynamic, asynchronous properties of the video, the model need to handle the short-term context and long-term context of the video story. The system learns the video with concept learning methods and there are two memories in which process observed video data of different time resolution. To generate questions from the video, the system has additional QA data for the video and learns the pattern of the image-question pairs. Generating question is similar to machine translation problem. To generate answers for the questions, it is needed to search both the short-term and long-term memory.

III. EXPERIMENTAL RESULTS

A. Cartoon Video and QA Dataset

Cartoon videos are a popular material for early language learning for children. They have a succinct and explicit story, which is represented with very simple images and easy words [3]. These properties allow the cartoon videos to be a test bed material suitable for a video question & answering played by a child and a robot [3]. For the experiment, we use a cartoon video ‘Pororo’ with 1200 question & answer pairs from ‘Pororo’ video. To preprocess the video, we convert the video into scene-subtitle pairs. Whenever the subtitle appears in the video, the scene at that time is captured. Each scene is converted to a set of image patches using Regions with Convolutional Neural Networks (R-CNN) [4], and the patch is represented by a 4096-dimensional fc7 feature of CNN using the Caffe implementation. In this work, we use Word2vec to encode the words

B. Network Construction

As the video stories unfold, the short-term memory and the long-term memory independently learn the observed video and thus show different emergence and evolution phases in the entire networks. Figure 2 shows that there are two area in the networks. The center area with red edges and nodes indicates the knowledge learned from the short-term memory and the outer area with gray edges and nodes indicates the knowledge learned from the long-term memory. The content of nodes related with the short-term memory continuously change during the learning process within the fixed area size while the content of nodes related with long-term memory rarely do not change within the expanded area size. We fixed the size of the short-term memory to 300 microcodes in this work.

C. Evaluation

To evaluate our model, we use our model to retrieve 80 answers from 80 questions. Then, twelve human evaluators judge the results whether the answers are given by a machine or a human. Each answer is presented with the question and a sequence of seven images related to the question in order to give information about the video story to the evaluators. Additionally, to test the expandability of our model, we apply the model to another cartoon video set ‘MaisyABC’ and evaluate the results. The results are shown in Table 1. It shows that 32.60% and 35.47% of the answers for each video are treated as answers from a human.

IV. CONCLUSION AND DISCUSSION

We proposed a deep learning-architecture of video question & answering for human-robot interaction. We demonstrate that the proposed architecture have both short-term and long-term contexts in the video and can generate questions and retrieve answers appropriately. We evaluate our methods using the real Turing Test. In future work, we plan to use much larger datasets including TV dramas and movies with more complex stories. Also, the system should be expanded to be a “purposive” or “intentional” agent. The system should be able to decide which observed data needs to be more focused on and which questions should be generated. These creative properties are necessary for lifelong learning environments [5].

Table 1: The Video Turing Test Results of the Retrieved Answers

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Pass</th>
<th>Fail</th>
<th>Pass Rate(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pororo</td>
<td>313</td>
<td>647</td>
<td>32.60</td>
</tr>
<tr>
<td>MaisyABC</td>
<td>127</td>
<td>231</td>
<td>35.47</td>
</tr>
</tbody>
</table>

REFERENCES

Abstract—Household robots will live with humans together, and then they should know general temporal knowledge of everyday lives in various time scales. For learning the temporal knowledge of family members, observation-interaction-oriented dataset is fundamental, but, such data to understand contextual stories are not available yet. As an alternative, one of available data for learning by showing to robots is series of cartoon videos for young kids. This type of data has some advantages: 1) omnibus style: simple and explicit storyline in short, 2) narrative order use fabula (chronological sequencing), 3) limited number of characters and spatial environment. Here, we introduce the framework to learn stories of cartoon videos. To represent stories, we define an event as the concatenation of a continuous vector from a scene description sentence and one of a dialogue sentence. So, we collected description sentences for visual scenes by persons, and try to embed event vector onto latent space with favor of consecutive events using ranking loss. Using a series of approximately 200 episodes of cartoon videos named ‘Pororo the Little Penguin’, we visualize trajectory-like embedded space. We expect this approach to achieve the following goals: 1) easy to interpret the episodic context, 2) easy to approximate multi-scale missing events, 3) easy to infer blank scenes from the videos.

Keywords—scene embedding; story learning; surrogate life data

I. INTRODUCTION

Recently, released are socially interactive household robots such as NAO, Pepper, and Jibo. In a few years, the robots will live humans together, and then they should know general knowledge of everyday lives including temporal knowledge in various time scales to understand human life patterns better. Since the knowledge of the family members is personalized and episodic, the robots should learn via observation and interaction in the environment. Ideal datasets for learning the temporal knowledge of family members are observation-interaction-oriented ones collected on real situated environments, but such data to understand contextual stories are not available in public yet. As an alternative, one of available data for learning by showing to robots is series of cartoon videos for young kids. They have some advantages: 1) omnibus style, which each episode has simple and explicit storyline in short, 2) narrative order mostly use fabula, which follows chronological sequencing of the events, whereas syuzhet is a term to designate the way a story is organized to enhance the effect of storytelling. 3) limited number of main characters and limited spatial environment. This is good for computational burden to need smaller complexity to learn. These properties are so desirable to provide the data similar to that of everyday lives in compact and explicit way.

In this paper, we assume the scenario as shown Figure 1: robots are watching cartoon videos on TV, and English subtitles are provided on the screen. That is, we pursue learning by showing to focus on story understanding.

Figure 1. Scenario: Robot Learning by showing video series. As simplified data, a video stream converted to the stream of snapshots of pairs of images and texts.

We built new dataset from 3D animation videos for kids, entitled ‘Pororo the Little Penguin’, consisting of 16,066 scene-dialogue pairs created from the video of 20.5 hours in total length, 27,328 fine-grained descriptive sentences for scene descriptions.

II. METHOD

A. Scene Event definition

Story itself is so various to handle, so we define a story as sequences of events. In this framework, an event is shortest element to represent a story. As in Figure 1, we captured snapshots of images and texts pairs, so we assume each snapshot has one event. And, the information in the images can be exposed with description sentences.
As a result, to represent stories, we define an event as the concatenation of a continuous vector from a scene description sentence and one of a dialogue sentence. We use "skip-thought vector[1]" to change continuous vectors from the sentences.

B. Embedding consecutive events

Recently, pair-wise ranking loss is used popular to learn association between multi-modal information to connect similar pairs. Mostly, hinge loss or triplet loss can be used. In this work, we used the following variant of triplet loss,

\[ \min \sum \sum \max(0, \alpha - s(x_t, x_{t+1}) + s(x_t, x_{t+1})) \]

\( x_t \) is one event vector at time index t, \( x_{t+1} \) is another event vector to be chosen randomly. This loss function drives event vector at \( t \) and \( t+1 \) close, arbitrary other vectors far.

III. EXPERIMENT AND VISUALIZATION

A. Caption generation

At first, to build generalized automated captioning module, we use neuraltalk2 built by Andrei Karpathy. It is composed of Convolutional neural networks and Long Short-term Memory(LSTM) to learn image and description sentence pairs. This tool also distribute pre-trained model with lots of data, but, if directly applied to our cases, it show bad results. So, we relearn our dataset on the models. The following figure shows some examples of the results.

![Figure 2. Caption generated with neuraltalk2](image)

B. Visualization

To embed event vectors with favor of consecutive events, we use Multi-layer perceptron (MLP) as a mapping function from event vectors to latent space. And we used the loss function in 2. B, and visualize the latent space changing the number of nodes in each layer of MLPs.

![Figure 3. Embedded event vector visualization with t-sne](image)

IV. Final goals

We expect this approach to achieve the following goals: 1) easy to interpret the episodic context, 2) easy to approximate multi-scale missing events, 3) easy to infer blank scenes from the videos.

REFERENCES


Abstract—Recent large-scale enterprise market required to provide a specific solution to the customers. In particular, most of company’s product is related to on-line based solution so that a mainstream of anomaly pattern is now investigated using a real measurement for a certain period. Anomaly pattern from the real data is illustrated by allocated other nodes, and objects (O1 and O2) are distinguished from the normal data. A strong link with the functionality of data processing is discussed and our series of research will be deliver in near future along with fraud detection system.

I. INTRODUCTION

Today’s large-scale enterprise market is required tremendous data transmissions through Internet based company, telephony based company, and On-line based company, such as Paypal, Alibaba, Amazon, etc. Most of companies are currently have an issue of product and system for customer care service, for example, insurance company need to provide some statement; automobile company also need to send a combined data – price and estimation – to existing and future customers. Often, abnormal pattern which is occurring while the customer’s data is not safely treated - not only security issue and also a not treated well - and then there is a significant damage both the company and the customers.

We are drowning in the deluge of data that are being collected world-wide, while starving for knowledge at the same time. Anomalous events occur relatively infrequently [1]. This paper analyzes the current position of fraud detection using data clustering associated with anomaly pattern mining and demonstrates an attempt to conduct a link on large-scale enterprise market.

In chapter 2 of this paper, defects on anomaly pattern structure of the traditional service and proposed configuration is narrated. In chapter 3 provides theoretical definition of fraud detection and clustering. In chapter 4 the solution from chapter 3 is proved for its fraud detection system. Lastly, in chapter 5 is the conclusion.

II. ANOMALY PATTERN PROBLEM

A. Definition of Anomaly

Anomaly is a pattern in the data that does not conform to the expected behaviour. Also referred to as outliers, exceptions, peculiarities, surprise, etc. Anomalies translate to significant (often critical) real life entities. One is Cyber intrusions, and the other is Credit card fraud. In real world anomalies are included, Credit Card Fraud (an abnormally high purchase made on a credit card) and cyber Intrusions (A web server involved in ftp traffic). Another example is credit card transactions, insurance claim fraud, opinion fraud, social security fraud, call behavior fraud, scientific fraud and even terrorism to protect national security [2].

B. Structure of Traditional Data Service

Traditional data service for large-scale enterprise market, there is a continuous feedback from an early stage of data clustering, data integration, pattern mining and fraud detection, throughout all traditional data service. Large-scale enterprise simply plays a role as fee charger and only access each step-by-step procedures. total power usage. Smart meter, however, is capable of other versatile uses. Figure 1 displays a feedback system that applies each step of data processing. In current enterprise market, their concrete involvement and assessment is data clustering and integration so that many global companies are provide many solution and products.

![Figure 1: Existing and (b) Proposed Network Configuration](image-url)

In the meanwhile, machine learning for working with social network data: detecting fraud, predicting click-thru patterns, targeted advertising, etc. Algorithm is involved with support vector machines (SVM), collaborative filtering, rule mining algorithms, and many others. The usage of machine learning and re-enforced learning mechanism is popular so that may researcher are applied to many areas. In large-scale enterprise is more focused on data processing, such as data gathering, clustering, and integration, etc.

Anomaly pattern is occurred in many place as well, even on-line market and insurance transactions. Most of cases are involved a step-by-step, data clustering, data integration and pattern mining, as shown in Fig.1. We also choose that a series of industry while we are investigating an On-line related
company, most likely energy-related company is first. Hence, we investigate the charging amount of electric vehicle (EV) in a certain period. In this case, our results are perfectly match with a traditional data analysis results. Figure-1 displays an anomaly pattern detection (#141, O₁) and O₂ rather than a normal N₁, N₂, and N₃. Also, there is a middle point (50%), an upper whisker (M.P. + 1.5 *(Q₃-Q₁)) and a lower whisker (M.P. + 1.5 *(Q₃-Q₁)). This is only achievable through real-time fare plan or one which can reflect the balance of demand-supply. It should be noted that N₁, N₂ and N₃ are regions of normal behaviour, the other Points O₁ and O₂ are anomalies. Points in region O₁ and O₂ are anomalies as shown in Fig 2. Fig 2 shows a grouped (regions of normal behaviour).

Figure 2: Anomaly Pattern from checking data #141, O₁ and O₂, not exclusive N₁, N₁, and N₃ data. (Actual Data Measurement and Analysis of EV Charging in May 2015)

III. FRAUD DETECTION ON GIVEN SYSTEM

To find fraud is not an easy task. Overall one of best approaches is Outlier detection, it can be defined as an abnormal behavior and/or characteristics in a data set might often indicate that that person perpetrates suspicious activities. In data set level, behavior of a person/instance does not comply with overall behavior. e.g., illegal set up of customer account. On the other hands, we can assume that there is another one which called, “data item level” [3]. Behavior of a person/instance does not comply with normal behavior of that person/instance. e.g., identity theft [4].

Also, it is a challenge because “fraud” is an uncommon, well-considered, imperceptibly concealed, time-evolving and often carefully organized crime which appears in many types and forms. First, “Uncommon” it is extremely skewed class distribution. Big data, but only few fraudulent observations (often < 1%). Hard for data mining algorithms to learn from Rebalancing techniques: oversampling, underdamping. Second, “Well-considered”, it is complex fraud structures are carefully planned. Outlier detection no longer sufficient: combination of patterns, preferably well-hidden. Historical changes in behavior: temporal weighting. Third, “Imperceptibly concealed”, it is subtlety of fraud: imitating normal behavior, even in identify theft. Fraud often first “sleeping”, pretending to be a good customer [5].

Fourth, “Time-evolving”, it is fraudsters learn from their mistakes and those of their predecessors [6]. Finally, “Carefully organized”, it is relationships between fraudsters because it is also fraudulent transactions often occur at the same (set of) merchants [7]. Enterprise market often inherit many suspicious resources from past fraudulent companies (social security fraud). Fraudsters tend to call/contact the same (set of) people once they stole someone’s identity [8].

IV. ANALYSIS OF FAULT DETECTION

The first industries to use data analysis techniques to prevent fraud were the telephony companies, the insurance companies and the banks [9]. One early example of successful implementation of data analysis techniques in the banking industry is the FICO. Fraud is an adaptive crime, so it needs special methods of intelligent data analysis to detect and prevent it. These methods exist in the areas of, data mining, machine learning and statistics.

Table 1: Functionality of Data Processing

<table>
<thead>
<tr>
<th>No</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Data Crawling</td>
</tr>
<tr>
<td>2</td>
<td>Data Clustering</td>
</tr>
<tr>
<td>3</td>
<td>Data Match/Merge</td>
</tr>
<tr>
<td>4</td>
<td>Data Integrity/Integration</td>
</tr>
<tr>
<td>5</td>
<td>Data Visualization</td>
</tr>
</tbody>
</table>

Table 1 shows a functionality of data processing, from 1 step (data crawling) to 5 step (data visualization).

V. CONCLUSION

This paper analyzed the anomaly pattern structure of traditional enterprise approaches and present a methodology for data clustering and integration throughout large-scale enterprise market. Also, it proposes a solution to resolve the fraud detection on given system and analysis.

ACKNOWLEDGEMENT

This research was supported by the MISP, Korea, under the National Program for Excellence in SW(R7719-16-1002) supervised by the IITP.

REFERENCES

Childbot: A Conversational Assistant for Childcare

Hwiyeol Jo
Department of Computer Science & Engineering
Seoul National University
Seoul, South Korea
hyjo@bi.snu.ac.kr

Dong-Sig Han
Department of Computer Science & Engineering
Seoul National University
Seoul, South Korea
dshan@bi.snu.ac.kr

Woo-Young Kang
Department of Computer Science & Engineering
Seoul National University
Seoul, South Korea
wykang@bi.snu.ac.kr

Byoung-Tak Zhang
Department of Computer Science & Engineering
Seoul National University
Seoul, South Korea
btzhang@bi.snu.ac.kr

Abstract— In this paper, we used the RNN Encoder-Decoder (seq2seq) model in Korean conversational robot (Konvbot). As a first step of being a general conversational model, we restricted our domain to childcare situation which includes six circumstances: 1) waking up, 2) morning exercise, 3) having breakfast, 4) taking a shower, 5) wearing clothes, and 6) going to school. We collected about 10,000 dialogue pairs for this scenario from more than 30 people. With the data, we implemented the base conversational model, which is used for collecting more dialogues in real environments (e.g. lab tour, conference demo). We present experiments in our expected scenario as well as general conversations which are out-of-script, and finally, real conversation with children. The result showed that our model could catch slight different expressions in the similar context, but it can cover only specific domain and has low vocabulary due to the small amount of training data. With more real experiments, we can collect data from experimenters, the better conversation the model will generate.

Keywords—chatbot; conversational model; seq2seq;

I. INTRODUCTION

Deep neural network models based on the recurrent neural networks (RNN) have shown great success in many application such as speech recognition [1] and natural language processing (NLP) tasks [2][3]. The RNN Encoder-Decoder (seq2seq) model has shown excellent performance in the machine translation task [4]. The seq2seq model has an advantage of sentence based machine translation, which is generating a target sentence from a source sentence directly. Not only that, research showed the seq2seq model can be used as a conversational model with little modification [5]. So, we adapted the seq2seq model in our Korean conversational robot whose goal is for childcare with a sentence to sentence conversations. After building our conversation model using the seq2seq, we embedded it into robot platform to interact with people directly through voices using the speech recognition technique and the text to speech technique. As a first step of being a general conversational model, we restricted our target domain task to childcare situation at home environment.

II. METHOD

A. Scenario

We restricted the global domain of natural language conversation problems to the situation of childcare at every morning, which is smaller and simpler (we expected), but commonly desired for dialogues with physically interacting robots. We categorized this scenario into six circumstances: 1) waking up, 2) morning exercise, 3) having breakfast, 4) taking a shower, 5) wearing clothes, and 6) going to the school. For qualitatively testing seamlessness of our implementation, we used an actual robot and a single room which was adorned like a home environment and attempted to talk with the robot in given scripts using a non-commercial Korean speech recognition and text-to-speech software.

B. Data

For collecting dialogue data for the scenario, more than 30 people produced synthetically generated sentences; each person was assigned to make 300 pairs of dialogues for the scenario using his/her imagination, and 300 instruction-operation pairs for constructing a personality of the robot and future research. Also, we made the data that the speaker in the first line and last line are always different, and the latter is for the model. We collected about 10,000 sentences pairs for the experiments. With the data, we implemented the base conversational model, which is used for collecting more dialogues in real environment. For example, we brought our model in lab tours and conferences, and got more dialogues from visitors, leaving conversation logs.

C. Sequence-to-Sequence Frameworks

Our approach is based on sequence-to-sequence (seq2seq) framework described in [5]. The model consists of two recurrent neural networks. The encoder network reads one word in the input sequence at a time and also predicts the output sequence one word at a time. It is trained by maximizing
the probability output sequence for estimating the true output in
the data.

Specifically, at the bottom, we have a word embedding
matrix for Korean words initialized arbitrarily, which could be
pre-trained word representation in the English setting. The
encoding LSTM (Long-Short Term Memory) [6] just receives
input words successively and does not produce any other
outputs except hidden states. The decoding network uses the
final hidden state of the encoder as its previous hidden values,
and also use its last output as current input. The probability of
output of the decoder is calculated by additional fully-
connected layer with softmax activation. Fig. 1 summarized
the architecture.

Fig. 1. Illustration of proposed architecture.

III. EXPERIMENT & RESULTS

A. Childcare Scenario

The followings are expected scenario in our childcare
domain. We present simplified version.

Bot: Good morning, wake up now.
Child: Good morning.
Bot: Yeah, did you sleep well?
Child: I couldn’t sleep well because of hot weather.
Bot: Then, shall we exercise?
Child: Sure.
Child: Wow, sausage! I love it.
Bot: Fried eggs are also there.
Child: Which clothes would I wear?
Bot: How about short pants? It’s hot in these days.
Child: Where is my football?
Bot: Let’s find together.
Bot: If finish to prepare, you should go to school.
Child: Goodbye.
Bot: Be careful.

B. General Conversation

After fitting the scenario, we tested our conversation
model in out-of-script conversation. Green color means
reasonable answer; Blue means ambiguous but not bad; Red
means it does not make sense.

Child: I’m hungry
Bot: Don’t you have lunch?
Child: Give me a dish

Bot: Won’t you have it by yourself? I’m tired.
Child: When will we have a meal?
Bot: Why not?
Child: How old are you?
Bot: None of your business!
Child: What is for lunch?
Bot: It is in closet
Child: Where is my glasses?
Bot: Isn’t it in your room?

The result showed that our model can catch slight different
expressions in the similar context, but it can cover only
specific domain and has small vocabulary size due to the small
amount of training data.

C. Real Conversation with Child

Child: Do you like a white radish?
Bot: No, give me more.
Child: Where are you come from?
Bot: Yes.
Child: Stand up.
Bot: In this way, I will fall.
Child: What is for lunch?
Bot: Just sitting
Child: You, stupid
Bot: Better than you
Child: What is 1+1?
Bot: What?

In this experiment, children asked unsuspected questions
to our model. Although our model tried to find the best
answer in its vocabulary, the answer was usually not good.

IV. CONCLUSION

In this paper, we proposed the conversational robot using
the Korean language in restricted domain. Using the RNN
Encoder-Decoder model (seq2seq) [5], we could implement
sentence based conversation model. We trained our model with
relatively small amount of data, so the model cannot generate
appropriate answers in many cases. However, with more real
experiments, we can collect data from experimenters, the better
conversation the model will generate.

REFERENCES

networks for large-vocabulary speech recognition.” IEEE Transactions
phrases and their compositionality." Advances in neural information
processing systems. 2013.
encoder-decoder for statistical machine translation.” arXiv preprint
[5] Sutskever, Ilya, Oriol Vinyals, and Quoc V. Le. “Sequence to sequence
learning with neural networks.” Advances in neural information
processing systems. 2014.
Human-machine Interactive Knowledge Discovery
- integrating IBM Watson Concept Insight in HCI and future research methodology

Hyunjeong Lee, User Experience Lab
Seoul National University
Seoul, Korea
hyunjeong.lee@snu.ac.kr

Joongseek Lee, User Experience Lab
Seoul National University
Seoul, Korea
joonlee8@snu.ac.kr

Abstract— Unstructured data is explosively increasing. Within this massive and growing mountain of unstructured data lies exceptional and transformational value, such as revealing social trends and insights. However, irregularities and ambiguities make it difficult to analyze, and the support of (semi) automatic analysis methods are yet unsatisfactory, especially in terms of discovering new knowledge. Therefore, in this research, we take upon the HCI-KDD approach with the main goal of supporting human intelligence with machine intelligence. By using the open source IBM Watson Concept Insights, we propose an interactive knowledge discovery research method, and conducted an empirical case study. We collected 259 articles, and extracted 1295 theme words to derive 15 future generation lifestyles. This research contributes to the empirical study of interactive knowledge discovery, and hopes to trigger more interest in future human-machine collaboration research methods.

Keywords— unstructured data analysis, knowledge discovery, HCI-KDD, research methodology

I. INTRODUCTION
Unstructured data is explosively increasing. Experts at Gartner, Forrester and IDC estimate that somewhere around 80 % of all the data in the world are not managed in a database. The best example of such unstructured data is text, including e-mail and blog posting to news articles and journal papers. Within this massive and growing mountain of unstructured data lies exceptional and transformational value [4], such as revealing social trends and insights. However, irregularities and ambiguities make text difficult to understand using traditional programs.

Many researchers have pointed out the next challenge in knowledge discovery from such data [2]. Although text is easily and frequently created, the support of (semi-) automatic analysis methods are yet unsatisfactory, especially in terms of discovering new knowledge. Contrary to the classical text mining, or information retrieval approach, where the goal is to find information, hence the researcher knows what he wants, in knowledge discovery we want to discover novel insights, get new knowledge which was previously unknown [2]. Therefore, rather than depending on pure computer science, the key lies in understanding human and computer interaction, communication, and sense-making while discovering new knowledge [6].

II. RELATED WORK
A. HCI-KDD Approach
The HCI-KDD approach [3] is a synergistic combination of methodologies and approach with the goal of supporting human intelligence with computational intelligence - by bringing the human into the loop. This approach appreciates both what humans can do best and what computers can do best. [3] Humans are good at pattern recognition, whereas computers are beneficial in dealing with a massive amount of data, in terms of speed and diversity due to the world wide web. Moreover, to design a machine learning algorithm such as artificial intelligence - needs a well defined goal and huge amounts of training data. Therefore, it is still difficult to generate undiscovered insights, fully depending on computers.

B. Data Analytics Software
QDAS such as NVivio and Atalas.ti have been researched from the 1980s [5]. However, such software mainly supports the manual coding process, therefore cannot be precisely be comprised as computational intelligence. More recently, there are open source software, programming languages, and packages within statistical analysis software support text analysis, such as IBM SPSS Modeler, SAS Text Analytics, and the tm package within R [2]. Other solutions utilize machine learning algorithms and techniques, with the most prominent frameworks Weka (Waikato Environment for Knowledge Analysis) and RapidMiner [2].

III. PROPOSING METHOD
In our project we are using IBM Watson Concept Insight, which has been currently integrated into IBM Watson Alchemy Language. By using this open-source software, we aimed to discover new knowledge about the ‘future lifestyle of millennials’. The methods are described below.
A. IBM Watson Concept Insight

WCI is an open service that extracts related concepts from the input document. WCI uses a built-in knowledge model, the concept graph that provides a scoring mechanism for how two concepts (or 2 sets of concepts) are related to each other that is computed using the personalized page rank algorithm (Page et al. 1999) [1].

B. Data Acquisition

We collected news articles which forecasted future lifestyles. It was important to select the proper size of query. The abstraction level of the query could influence the data acquired. If so, the data that is acquired already possess a specific type of knowledge, whereas the objective of this methodology is to find new insights. In this case, ‘lifestyle’ was too broad and abstract, so we had to breakdown the scope. Noting to Nielsen lifestyle reports, we used work, eat, play, live, eat as queries instead of lifestyle. We collected 50 articles from each sub-category, and a total of 259 articles were collected. The title, sub-title, and body of the text were included in the analysis, and the advertisements were excluded.

C. Data Analysis

The data mining was rather simple and cost efficient, hence WCI did not require additional computer programming. The text from the news articles were copied and pasted into the ‘body of text’ of the WCI interface (fig.2). Only the top 3 related concepts were exposed, however by the ‘view JSON panel’ (fig.2) we could obtain a total of 5 related concepts. Consequently, a total of 1295 theme words were extracted. These key words were organized on excel (fig. 1), having the 5 theme words listed horizontally upon each article. Then each researcher highlighted the keywords upon a ‘deducting protocol’, since not all of the resulting keywords were valuable.

Keywords that contained adjectives or newly-coined words were considered most valuable. Then researchers used the affinity diagram (Beyer and Holtzblatt 1997) method using the selected key words. These process is considered as the ‘human-in-the-loop’ where human intelligence interacts with computational intelligence (extracted key words from WCI) to discover new knowledge and meaning from unstructured data.

D. Results

Researchers found three lifestyle concepts for each sub-category lifestyle: work, live, eat, play, save. The results are shown in Table1.

<table>
<thead>
<tr>
<th>Concept</th>
<th>No. 1</th>
<th>No. 2</th>
<th>No. 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>work</td>
<td>Playbour: The new workaholism</td>
<td>Tech supports freelancers</td>
<td>Social Media as portfolio</td>
</tr>
<tr>
<td>live</td>
<td>Co-habitation as a strategy</td>
<td>What I can reach over what I have</td>
<td>Gender as a spectrum</td>
</tr>
<tr>
<td>eat</td>
<td>Healthy fast-food</td>
<td>Drinking functional ingredients</td>
<td>Kitchen does the cooking</td>
</tr>
<tr>
<td>play</td>
<td>+ Mental health</td>
<td>AR/VR at the door</td>
<td>Live-on video</td>
</tr>
<tr>
<td>save</td>
<td>Now and me</td>
<td>Pay to play and upload</td>
<td>Shopping for experience</td>
</tr>
</tbody>
</table>

IV. CONCLUSION

This research proposes a novel research methodology, which inquires interactive knowledge discovery between human and machine, by auto-extracting concepts from unstructured text data. Also by conducting the actual application of the methodology, contributes to the empirical study of interactive knowledge discovery, and hopes to trigger more interest in future human-machine collaboration research methods. The limitation of this research is that the cognitive aspects of the user, has not yet been thoroughly scrutinized. Therefore, we further plan to conduct user experience evaluation and research, regarding interactive knowledge discovery.

REFERENCES

Beyond AlphaGo