Computing and Information Systems

4TH ANNUAL DOCTORAL COLLOQUIUM

Program and Proceedings

13th July 2016
Welcome to the CIS 2016 Doctoral Colloquium!

For the fourth consecutive year, The Department of Computing & Information Systems (CIS) at The University of Melbourne, is hosting the Doctoral Colloquium (DC) and on behalf of the 4th CIS DC organizing committee, we are honoured and delighted to welcome each of you to this annual event.

The CIS DC aims to create opportunities for graduate research students by motivating them to explore new research areas and foster their awareness of recent advances and updates offered by other students. The colloquium targets a broad range of participants, from new candidates to those who are ready to submit their thesis. By providing a collaborative and interactive environment bringing together academics, industry partners and students, all participants are able to benefit from their involvement in this event. Paper presentations and poster sessions are included in the program to augment the essential social and presentation skills of our graduate research students.

We have an exciting Q & A session planned this year, “Life After PhD”, to take a closer look at potential challenges a PhD graduate may face. A distinguished panel with academic, research and industry backgrounds will share their experience with students and answer questions.

The success of the CIS DC 2016 is due to the hard work of staff and graduate research students in CIS who have worked with us in planning and organizing both the technical program and supporting social arrangements. In particular, we would like to thank Justin Zobel, Jan Schroeder, Rhonda Smithies and Julie Ireland for their much valued advice and support. Needless to say, this colloquium could not happen without the involvement of our graduate research students who have volunteered their time by serving on the organising committee and we would like to acknowledge their hard work, energy and passion for this event. Finally, we would like to thank our sponsors, Australian Computer Society, Google, Microsoft, Victorian Life Sciences Computation Institute, Melbourne Networked Society Institute, the Microsoft Research Centre for Social Natural User Interfaces, the Melbourne School of Information and the Melbourne School of Engineering, for their support of the colloquium.

We hope that you will find attending this year’s CIS DC a rewarding experience.

CIS DC 2016 Organising Committee
Conference Location

Herbert Wilson Theatre
Ground Level,
Doug McDonell Building (193)
Campus Map reference: J 19

Seminar Rooms 103 and 202
Old Metallurgy Building (166)
Wilson Avenue
Campus Map reference: J 17

Registration and Keynote

Paper, Poster and Q&A Sessions

Map of Conference Venue

Source: http://lostoncampus.com.au
Source: http://learningspaces.unimelb.edu.au
Conference Program

8:30 - 9:00
Registration
Herbert Wilson Theatre, Doug McDonell Building

9:00
Opening and Welcome Address
Prof Justin Zobel
Herbert Wilson Theatre, Doug McDonell Building

Keynote:
From digital to neural disruption: the role of intelligent machines in society - a financial services case study
Drs Patrick Maes
Herbert Wilson Theatre, Doug McDonell Building

9:10 - 10:00
10:00 - 10:20
Morning Tea Break
Level 2, Old Metallurgy Building

Session 1A
Data and Knowledge
Old Metallurgy-103
Unsupervised Parameter Estimation for OCSVMs
Zahra Ghafoori
Improving electricity consumption prediction in a campus environment
Pasan Karunaratne
Investigating the Behaviour of Word Vector Models on Smaller Datasets
Sameendra Samarawickrama
An Efficient Algorithm for Non-Interactive Differential Privacy
Soheila Ghane
A Bayesian Approach for Inferring the Hidden Truth from Crowdsourced Annotations
Yuan Li

Session 1B
People and Organisations
Old Metallurgy-202
Observing boardgame play in a research setting
Melissa Rogerson
The immobility of mobile devices: insights from video-mediated sessions with ethnic communities in Kenya
Kagonya Awori
Mapping Melbourne Children’s Digital Play
Jane Mavoa
Designing Avatars to Promote Social Engagement for Older Adults
Romina Carrasco
The Effect of Gaze on Gameplay in Co-located Multiplayer Gaming Environments
Joshua Newn
<table>
<thead>
<tr>
<th>Time</th>
<th>Session 2A Platforms and Systems</th>
<th>Session 2B People and Organisations</th>
</tr>
</thead>
<tbody>
<tr>
<td>11:20 - 12:20</td>
<td>Mitigating Impact of Short-term Overload on Multi-Cloud Web Applications through Geographical Load Balancing Chenhao Qu</td>
<td>Exploring information artefacts of person centred discharge planning Nyree Taylor</td>
</tr>
<tr>
<td></td>
<td>Energy-efficient Scheduling of Cloud Application Components with Brownout Minxian Xu</td>
<td>Increasing Clinical Data Entry Quality Through Automated Web-based Feedback Stephan Glöckner</td>
</tr>
<tr>
<td></td>
<td>Towards Performance-Oriented Deployment of Streaming Applications on Cloud Xunyun Liu</td>
<td>Technology at Mealtimes: The Good, the Bad, and the Ugly Hasan Shahid Ferdous</td>
</tr>
<tr>
<td></td>
<td>SLA-aware and Energy-efficient Dynamic Overbooking in SDN-Enabled Cloud Data Centers Jungmin Son</td>
<td>Some Things Change, Some Stay the Same: The Impact of Book Format on Borrowing Patterns Dana Mckay</td>
</tr>
<tr>
<td></td>
<td>Semantic-aware Service Definition and Discovery in IoT Using Linked Data and CoAP Farzad Khodadadi</td>
<td>A Learner Centric Approach to a Knowledge Sharing Virtual Community for General Practitioners Abdulaziz Murad</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time</th>
<th>Mid-Day Break (Lunch Not Provided)</th>
</tr>
</thead>
<tbody>
<tr>
<td>12:20 - 13:10</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time</th>
<th>Session 3A Data and Knowledge</th>
<th>Session 3B Data and Knowledge</th>
</tr>
</thead>
<tbody>
<tr>
<td>13:10 - 14:00</td>
<td>A segmentation method for counting repeats in a weightlifting routine Yousef Kowsar</td>
<td>Exploiting Tree Kernels for high performance Chemical Induced Disease relation extraction Nagesh Panyam Chandrasekarasstry</td>
</tr>
<tr>
<td></td>
<td>Dear Noise: Hime Me Please! Elham Naghizade</td>
<td>Reproducibility of Genomic Workflows Sehrish Kanwal</td>
</tr>
<tr>
<td></td>
<td>Biographical Fact Extraction using non-linear SVM kernels Gitansh Khirbat</td>
<td>Machine-Learning Algorithms Predict Graft Failure Following Liver Transplantation Yamuna Kankanige</td>
</tr>
<tr>
<td></td>
<td>Distributed Coding of time series data Lakshmi J Mohan</td>
<td>Semantic-based Privacy-preserving Record Linkage Yang Lu</td>
</tr>
</tbody>
</table>
### Session 4A
**Platforms and Systems**
*Old Metallurgy-103*

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
</tr>
</thead>
</table>
| 14:00 - 15:15 | Finding Data Races in Java programs  
David Clarke |
|            | Parallel Algorithms for k-Center Clustering  
Jessica McClintock |
|            | Privacy Protection of Large-scale Trajectories  
Shuo Wang |
|            | Constraint Programming Approach to Find Bounded-Distance Paths in a Graph  
Diego de Uña |
|            | Effective Straggler Mitigation for Scalable Distributed Microscopic Traffic Simulation  
Eman Bin Khunayn |
|            | Combining BDD based Circuit Synthesis Technique with Masked Dual-Rail Pre-charge Logic to Eliminate Glitches in Circuits  
Partha De |

### Session 4B
**People and Organisations**
*Old Metallurgy-202*

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
</tr>
</thead>
</table>
| 14:00 - 15:15 | Information Security Management Practices in Organisations  
Moneer Alshaikh |
|            | Information Security: A Short Essay on a Theory  
Craig A. Horne |
|            | Business Analytics In Information Security Risk Management: The Contingent Effect On Security Performance  
Humza Naseer |
|            | Mitigating Knowledge Leakage Risk in Organizations through Mobile Devices: A Contextual Approach  
Carlos Andres Agudelo Serna |
|            | Understanding How Start-ups Gain a Competitive Advantage from Cloud Computing  
Mohammed Alrokayan |
|            | Information Security Strategy: Beyond Protection, Towards an Organisational Information Security Capability  
Mazino Onibere |

### Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
</table>
| 15:15 - 15:30 | Afternoon Tea Break  
*Level 2, Old Metallurgy Building* |
| 15:30 - 16:30 | Q&A Panel  
*Life after PhD*  
*Old Metallurgy-202* |
| 16:30 - 17:30 | Poster Presentations  
(End of the day refreshments)  
*Old Metallurgy-205* |
| 17:30 - 18:00 | Closing and Award Ceremony  
*Foyer Area, Level 2, Old Metallurgy Building* |
What do you think about that?

Tweet us @cis_dc and share your thoughts on the talks / presentations / posters you are listening to!

The hashtag to use is #cisdc2016

Find us on facebook:
https://www.facebook.com/melbcisdc
and give your valuable feedback about talks / presentations / posters you are listening to!
From digital to neural disruption: the role of intelligent machines in society - a financial services case study

The use of artificial intelligence, particularly in the financial services industry, has received much exposure recently. Yet it is by no means new.

This presentation will outline perspectives on AI and cognitive computing, its application, and implications for the workforce in a financial services context.

Bio Sketch

Patrick is an executive with over 30 years’ experience in banking, strategy, transformation, advanced technology, architecture, systems development, and management consulting.

Patrick is currently General Manager, Strategy & Planning for TSO - ANZ’s technology, shared services, operations, and property division. He is responsible for defining the strategy and plan of the division, and driving digital transformation initiatives for ANZ.

He has worked with many companies across Europe, Asia, Australia and North America including HCL Technologies, Bankers Trust Alex Brown, start-up investment bank WIT Capital, Rabobank International, IBM, and AXA amongst others.

Patrick has a Master in Commercial and Financial Sciences and an Aggregate Diploma for Higher Education - both from the Catholic University of Louvain. He also completed a Doctoral Program (Drs.) in Applied Economics specialising in Computer Sciences from the University of Antwerp, and a Postgraduate in Knowledge Technology, Artificial Intelligence, and Cognitive Sciences at the University of Ghent.

He serves on the advisory board of Melbourne’s RMIT School of Computer Science and Information Technology (SCSIT), is an Industry Advisory Group Member (Dept. of Computing and Information Systems) at the University of Melbourne, and is an Honorary Enterprise Professor also at the University of Melbourne.

Married with five children, his other interests include cooking, gardening, and fly fishing when he can make time.
## Paper Submissions

### Data and Knowledge

<table>
<thead>
<tr>
<th>Title</th>
<th>Authors</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exploiting Tree Kernels for high performance Chemical Induced Disease relation extraction</td>
<td>Nagesh Panyam Chandrasekarasatry, Rao Kotagiri, Karin Verspoor and Trevor Cohn</td>
<td>15</td>
</tr>
<tr>
<td>Unsupervised Parameter Estimation for OCSVMs</td>
<td>Zahra Ghafoori, Sutharshan Rajasegaran, Sarah Erfani, Shanika Karunasekera and Christopher Leckie</td>
<td>16</td>
</tr>
<tr>
<td>Reproducibility of Genomic Workflows</td>
<td>Sehrish Kanwal, Andrew Lonie, Richard Sinnott and Farah Khan</td>
<td>17</td>
</tr>
<tr>
<td>Improving electricity consumption prediction in a campus environment</td>
<td>Pasan Karunarantne, Shanika Karunasekera, Aaron Harwood and Masud Moshtaghi</td>
<td>18</td>
</tr>
<tr>
<td>Investigating the Behaviour of Word Vector Models on Smaller Datasets</td>
<td>Sameendra Samarawickrama, Shanika Karunasekera, Aaron Harwood and Rao Kotagiri</td>
<td>19</td>
</tr>
<tr>
<td>Machine-Learning Algorithms Predict Graft Failure Following Liver Transplantation</td>
<td>Yamuna Kankanige, Lawrence Lau, Benjamin Rubinstein, Robert Jones, Christopher Christphi, James Bailey and Vijayaragavan Muralidharan</td>
<td>20</td>
</tr>
<tr>
<td>Semantic-based Privacy-preserving Record Linkage</td>
<td>Yang Lu and Richard Sinnott</td>
<td>21</td>
</tr>
<tr>
<td>A segmentation method for counting repeats in a weightlifting routine</td>
<td>Yousef Kowsar, Eduardo Velloso, Lars Kulik, Christopher Leckie and Masud Moshtaghi</td>
<td>22</td>
</tr>
<tr>
<td>Dear Noise: Hime Me Please!</td>
<td>Elham Naghizade, James Bailey, Lars Kulik and Egemen Tanin</td>
<td>23</td>
</tr>
<tr>
<td>Supervised Learning Meets Biocuration: Detecting Duplicate Records in Biological Databases</td>
<td>Qingyu Chen, Justin Zobel and Karin Verspoor</td>
<td>24</td>
</tr>
<tr>
<td>Automatic Detection of Choroid-Sclera Interface in EDI-OCT Images</td>
<td>Md Akter Hussain, Alauddin Bhuiyan and Kotagiri Ramamohanrao</td>
<td>25</td>
</tr>
<tr>
<td>The Hitchhiker’s Guide to the Optimal Route Planning</td>
<td>Oleksii Vedernikov</td>
<td>26</td>
</tr>
<tr>
<td>An Efficient Algorithm for Non-Interactive Differential Privacy</td>
<td>Soheila Ghane, Lars Kulik and Ramamohanrao Kotagiri</td>
<td>27</td>
</tr>
<tr>
<td>Biographical Fact Extraction using non-linear SVM kernels</td>
<td>Gitansh Khirbat, Rui Zhang and Jianzhong Qi</td>
<td>28</td>
</tr>
<tr>
<td>A Bayesian Approach for Inferring the Hidden Truth from Crowdsourced Annotations</td>
<td>Yuan Li, Meng Fang, Benjamin I. P. Rubinstein and Trevor Cohn</td>
<td>29</td>
</tr>
<tr>
<td>Distributed Coding of time series data</td>
<td>Lakshmi J Mohan, Udaya Parampalli and Aaron Harwood</td>
<td>30</td>
</tr>
</tbody>
</table>

### Platforms and Systems

<table>
<thead>
<tr>
<th>Title</th>
<th>Authors</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Finding Data Races in Java programs</td>
<td>David Clarke, Antonette Mendoza and Tim Miller</td>
<td>32</td>
</tr>
<tr>
<td>Parallel Algorithms for k-Center Clustering</td>
<td>Jessica Mcclintock</td>
<td>33</td>
</tr>
<tr>
<td>Privacy Protection of Large-scale Trajectories</td>
<td>Shuo Wang and Richard Sinnott</td>
<td>34</td>
</tr>
<tr>
<td>Constraint Programming Approach to Find Bounded-Distance Paths in a Graph</td>
<td>Diego de Uña, Graeme Gange, Peter Schachte and Peter J. Stuckey</td>
<td>35</td>
</tr>
<tr>
<td>Authors</td>
<td>Title</td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>-------</td>
<td></td>
</tr>
<tr>
<td>Eman Bin Khunayn, Shanika Karunasekera, Kotagiri Ramamohanarao and Hairoof Xie</td>
<td>Effective Straggler Mitigation for Scalable Distributed Microscopic Traffic Simulation</td>
<td></td>
</tr>
<tr>
<td>Chenhao Qu, Rodrigo Calheiros and Rajkumar Buyya</td>
<td>Mitigating Impact of Short-term Overload on Multi-Cloud Web Applications through Geographical Load Balancing</td>
<td></td>
</tr>
<tr>
<td>Safiollah Heidari and Rajkumar Buyya</td>
<td>Cost-efficient Processing of Large-scale Graphs on Public Clouds</td>
<td></td>
</tr>
<tr>
<td>Minxian Xu, Amir Vahid Dastjerdi and Rajkumar Buyya</td>
<td>Energy-efficient Scheduling of Cloud Application Components with Brownout</td>
<td></td>
</tr>
<tr>
<td>Xunyun Liu, Rodrigo Calheiros and Rajkumar Buyya</td>
<td>Towards Performance-Oriented Deployment of Streaming Applications on Cloud</td>
<td></td>
</tr>
<tr>
<td>Neelofar Neelofar, Lee Nais and Rao Kotagiri</td>
<td>Improving Spectral Based Fault Localization Using Simple Static Analysis</td>
<td></td>
</tr>
<tr>
<td>Partha De, Udaya Parampalli and Chittaranjan Mandal</td>
<td>Combining BDD based Circuit Synthesis Technique with Masked Dual-Rail Pre-charge Logic to Eliminate Glitches in Circuits</td>
<td></td>
</tr>
<tr>
<td>Jungmin Son and Rajkumar Buyya</td>
<td>SLA-aware and Energy-efficient Dynamic Overbooking in SDN-Enabled Cloud Data Centers</td>
<td></td>
</tr>
<tr>
<td>Farzad Khodadadi</td>
<td>Semantic-aware Service Definition and Discovery in IoT Using Linked Data and CoAP</td>
<td></td>
</tr>
<tr>
<td>Bowen Zhou and Rajkumar Buyya</td>
<td>An Online Algorithm for Task Offloading in Heterogeneous Mobile Clouds</td>
<td></td>
</tr>
</tbody>
</table>

**People and Organisations**

<table>
<thead>
<tr>
<th>Authors</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nyree Taylor, Reeva Lederman and Rachelle Bosua</td>
<td>Exploring information artefacts of person centred discharge planning</td>
</tr>
<tr>
<td>Stephan Glöckner and Richard Sinnott</td>
<td>Increasing Clinical Data Entry Quality Through Automated Web-based Feedback</td>
</tr>
<tr>
<td>Deepti Aggarwal</td>
<td>Understanding the Role of Bodily Communication in Video Consultations for Physiotherapy</td>
</tr>
<tr>
<td>Melissa Rogerson, Martin Gibbs and Wally Smith</td>
<td>Observing boardgame play in a research setting</td>
</tr>
<tr>
<td>Hasan Shahid Ferdous</td>
<td>Technology at Mealtimes: The Good, the Bad, and the Ugly</td>
</tr>
<tr>
<td>Moneer Alshaikh, Sean B. Maynard, Atif Ahmad and Shanton Chang</td>
<td>Information Security Management Practices in Organisations</td>
</tr>
<tr>
<td>Kagonya Awori</td>
<td>The immobility of mobile devices: insights from video-mediated sessions with ethnic communities in Kenya</td>
</tr>
<tr>
<td>Fernando Estrada</td>
<td>Gay men’s perceptions of mobile apps for mental health</td>
</tr>
<tr>
<td>Jane Mavoa, Martin Gibbs and Marcus Carter</td>
<td>Mapping Melbourne Children’s Digital Play</td>
</tr>
<tr>
<td>Patricia Pons, Javier Jaen, Marcus Carter and Sarah Webber</td>
<td>Tangible Sound-Based Enrichment for Orang-utans</td>
</tr>
<tr>
<td>Romina Carrasco</td>
<td>Designing Avatars to Promote Social Engagement for Older Adults</td>
</tr>
<tr>
<td>Craig A. Horne</td>
<td>Information Security: A Short Essay on a Theory</td>
</tr>
<tr>
<td>Humza Naseer</td>
<td>Business Analytics in Information Security Risk Management: The Contingent Effect on Security Performance</td>
</tr>
<tr>
<td>Dana Mckay</td>
<td>Some Things Change, Some Stay the Same: The Impact of Book Format on Borrowing Patterns</td>
</tr>
<tr>
<td>Authors</td>
<td>Title</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Carlos Andres Agudelo Serna, Rachelle Bosua, Sean B Maynard and Atif Ahmad</td>
<td>Mitigating Knowledge Leakage Risk in Organizations through Mobile Devices: A Contextual Approach</td>
</tr>
<tr>
<td>Diana Wong, Rachelle Bosua, Shanton Chang and Sherah Kurnia</td>
<td>Individual Use of Enterprise 2.0</td>
</tr>
<tr>
<td>Mohammad Moniruzzaman, Sherah Kurnia, Sean B. Maynard and Alison Parkes</td>
<td>Business Intelligence and Supply Chain Agility</td>
</tr>
<tr>
<td>Mohammed Alrokayan, Wally Smith and Sherah Kurnia</td>
<td>Understanding How Start-ups Gain a Competitive Advantage from Cloud Computing</td>
</tr>
<tr>
<td>Mazino Onibere, Atif Ahmad and Sean Maynard</td>
<td>Information Security Strategy: Beyond Protection, Towards an Organisational Information Security Capability</td>
</tr>
<tr>
<td>Joshua Newn</td>
<td>The Effect of Gaze on Gameplay in Co-located Multiplayer Gaming Environments</td>
</tr>
<tr>
<td>Abdulaziz Murad</td>
<td>A Learner Centric Approach to a Knowledge Sharing Virtual Community for General Practitioners</td>
</tr>
</tbody>
</table>
Poster Submissions

**Data and Knowledge**

<table>
<thead>
<tr>
<th>Authors</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yu Sun, Rui Zhang, Andy Yuan Xue, Jianzhong Qi and Xiaoyong Du</td>
<td>Reverse Nearest Neighbor Heat Maps: A Tool for Influence Exploration</td>
</tr>
<tr>
<td>Piyapong Khumrin, Anna Ryan, Terry Judd and Karin Verspoor</td>
<td>The use of clinical decision support systems for the development of medical students’ diagnostic reasoning skills</td>
</tr>
<tr>
<td>Yu Sun</td>
<td>Contextual Intent Tracking for Personal Assistants</td>
</tr>
<tr>
<td>Abdullah Aldwyish, Egemen Tanin and Shanika Karunasekera</td>
<td>Location-based social networking for obtaining personalised driving advice</td>
</tr>
<tr>
<td>Xiaojie Lin</td>
<td>Linking News to Tweets</td>
</tr>
<tr>
<td>Hengfeng Li, Lars Kulik and Rao Kotagiri</td>
<td>Automatic Generation and Validation of Road Maps From GPS Trajectory Data Sets</td>
</tr>
<tr>
<td>Yiqing Zhang, Rui Zhang, Yuan Li and He Zhao</td>
<td>Detecting and Recognizing Text from Natural Scene Images</td>
</tr>
<tr>
<td>Afshin Rahimi, Trevor Cohn and Timothy Baldwin</td>
<td>Geolocation of Social Media Users</td>
</tr>
<tr>
<td>Yitong Li, Trevor Cohn and Timothy Baldwin</td>
<td>Learning Robust Representations of Text</td>
</tr>
<tr>
<td>Qingyu Chen, Justin Zobel and Karin Verspoor</td>
<td>Benchmarks for duplicate detection methods in biological databases</td>
</tr>
<tr>
<td>Xin Li, Kathleen Gray and Karin Verspoor</td>
<td>Using Learning Analytics to Understand Informal Learning of Health Professionals in Online Social Networks</td>
</tr>
<tr>
<td>Miji Choi, Karin Verspoor and Justin Zobel</td>
<td>Impact of Coreference Resolution on Biological Event Extraction</td>
</tr>
<tr>
<td>Tuan Doan and Christopher Leckie</td>
<td>Impact of Night Network on Pedestrian Activities</td>
</tr>
<tr>
<td>Aref Eshghishargh, Kathleen Gray, Simon Milton, Scott Kolbe and Andrew Lonie</td>
<td>Semantic Question Classification in Neuroscience by Using the Domain Knowledge</td>
</tr>
<tr>
<td>Milad Chenaghlou, Moshtagh Masud and Christopher Leckie</td>
<td>Anomaly Detection in Data Streams</td>
</tr>
<tr>
<td>Oleksii Vedernikov</td>
<td>The Hitchhiker’s Guide to the Pick-Up Locations</td>
</tr>
<tr>
<td>Sadia Nawaz, James Bailey and Gregor Kennedy</td>
<td>Early Detection of Confusion in Digital Learning Environments</td>
</tr>
<tr>
<td>Tan Ping Ping, Karin Verspoor &amp; Tim Miller</td>
<td>Versioned Text Documents Pre-Processing</td>
</tr>
<tr>
<td>Liyan Liu</td>
<td>Image Processing Pipeline in Time Lapse Fluorescence Microscopy</td>
</tr>
</tbody>
</table>

**Platforms and Systems**

<table>
<thead>
<tr>
<th>Authors</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leyla Roohi, Vanessa Teague and Udaya Parampalli</td>
<td>Privacy-Preserving Computations in Australian Metadata</td>
</tr>
<tr>
<td>Ronal Singh, Liz Sonenberg and Tim Miller</td>
<td>Shared Mental Models for Artificial Agent Teams Engaged in Interdependent Tasks</td>
</tr>
<tr>
<td>Anthony Quattrone, Lars Kulik and Egemen Tanin</td>
<td>Combining Range-Based and Range-Free Methods: A Unified Approach for Localization</td>
</tr>
<tr>
<td>Safiollah Heidari and Rajkumar Buyya</td>
<td>Cost-efficient Processing of Large-scale Graphs on Public Clouds</td>
</tr>
<tr>
<td>Yude Lin</td>
<td>Forward, Backward, and Compositional Symbolic Execution</td>
</tr>
<tr>
<td>Authors</td>
<td>Title</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Md Redowan Mahmud, Rao Kotagiri and Rajkumar Buyya</td>
<td>Fog Computing: Integrating IoT Sensors and Clouds</td>
</tr>
<tr>
<td>Sara Kardani Moghaddam, Rodrigo Calheiros, Rajkumar Buyya and Rao Kotagiri</td>
<td>Cloud Resource Management based on Workload Analysis and Anomaly Prediction</td>
</tr>
<tr>
<td>Friedrich Burkhard von der Osten, Michael Kirley and Tim Miller</td>
<td>The minds of many</td>
</tr>
<tr>
<td>Deepti Aggarwal</td>
<td>Supporting Bodily Communication in Video Consultations for Physiotherapy</td>
</tr>
<tr>
<td>Sarah Webber</td>
<td>Designing technology for use at the zoo</td>
</tr>
<tr>
<td>Jongkil Jeong</td>
<td>The Psychological Contract and ICT Outsourcing</td>
</tr>
<tr>
<td>Fernando Estrada</td>
<td>‘I am warmer than a machine and I listen’:Clinicians’ opinions of mobile apps for mental health</td>
</tr>
<tr>
<td>Haifa Binsahl</td>
<td>Exploring Saudi female international students’ information needs and seeking behaviour when moving across two different digital and cultural environments</td>
</tr>
<tr>
<td>Marion Zalk and Rachelle Bosua</td>
<td>Factors influencing Knowledge Transfer about Complex Tasks</td>
</tr>
<tr>
<td>Patricia Pons</td>
<td>Tangible User Interfaces for Animals</td>
</tr>
<tr>
<td>Toni Mclennan</td>
<td>Social Media and Self-concept: Does social media change the way we understand our self?</td>
</tr>
<tr>
<td>Nwakego Isika, Antonette Mendoza and Rachelle Bosua</td>
<td>Appropriation of Social Media by Adults with chronic illness</td>
</tr>
</tbody>
</table>
Data and Knowledge
Exploiting Tree Kernels for high performance Chemical Induced Disease relation extraction

[Extended Abstract]

P.C Nagesh
University of Melbourne, Melbourne, Australia
npanyam@student.unimelb.edu.au

ABSTRACT
In this paper, we study the effectiveness of tree kernels for Chemical-Disease relation extraction. We demonstrate that subset tree kernels increase the F-score to 61.7% as compared to 57.9% achieved with simple feature engineering.

CCS Concepts
• Computing methodologies → Kernel methods; Information extraction; • Applied computing → Bioinformatics;

Keywords
Syntactic Tree Kernels; Biomedical relation extraction

1. INTRODUCTION
The recent shared task for Chemical-Induced-Disease(CID) relation extraction organized by BioCreative-V (Wei et al., 2015), has made available a large body of annotated PubMed abstracts for the valuable Chemical-Disease relations. The task showed that supervised classification with a rich feature set can lead to high performance relation extraction. Designing such an effective feature set involves extensive feature engineering and is likely to be a task specific and almost certainly a domain dependent activity. Kernel methods in NLP (Collins & Duffy, 2001) have been designed precisely to address this problem of manual feature engineering. They enable an efficient and comprehensive exploration of a very high dimensional feature space and to automatically adapt to the dominant patterns in the training set.

In our work, we show that kernel methods can be used for boosting relation extraction performance without having to manually engineer a rich feature set. We demonstrate through experiments that combining tree kernels over constituent parse trees of input sentences.

3. TREE KERNELS
The syntactic structure of a sentence can be represented by its constituent parse tree and the syntactic similarity between two sentences can be computed using a tree kernel over its syntactic trees. A tree kernel implicitly maps a given tree into an exponential feature space of tree fragments(subtrees or subset trees). The unnormalized score is expressed as: \[ \sum_{n_1 \in N_T_1} \sum_{n_2 \in N_T_2} \Delta(n_1, n_2) \] where \( N_T_1 \) and \( N_T_2 \) are the sets of nodes of trees \( T_1 \) and \( T_2 \) respectively and \( \Delta(n_1, n_2) = \sum_{i=1}^{\left|F_i\right|} I_i(n_1)I_i(n_2) \). \( I_i(n) \) is an indicator function that evaluates to 1 if the fragment \( f_i \) is rooted at node \( n \) and 0 otherwise.

4. RESULTS
Our experiments show that tree kernels are very effective for intrasentence relation extraction (see Table 1).

<table>
<thead>
<tr>
<th>Test Data</th>
<th>Classifier</th>
<th>P</th>
<th>R</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intrasentence</td>
<td>Lin</td>
<td>58.2</td>
<td>75.6</td>
<td>65.8</td>
</tr>
<tr>
<td>Intrasentence</td>
<td>Lin + TK</td>
<td>63.3</td>
<td>75.4</td>
<td>68.8</td>
</tr>
<tr>
<td>Intersentence</td>
<td>Lin</td>
<td>33.7</td>
<td>39.8</td>
<td>36.5</td>
</tr>
<tr>
<td>Intersentence</td>
<td>Lin + TK</td>
<td>65.9</td>
<td>20.0</td>
<td>30.8</td>
</tr>
<tr>
<td>Full test</td>
<td>Lin</td>
<td>57.8</td>
<td>65.6</td>
<td>57.9</td>
</tr>
<tr>
<td>Full test</td>
<td>Lin + TK</td>
<td>63.6</td>
<td>59.8</td>
<td>61.7</td>
</tr>
</tbody>
</table>

Table 1: Results on CID test data for Linear classifier(Lin), Tree Kernel(TK) and their combination(Lin+TK).

5. SUMMARY AND FUTURE WORK
We have shown that intrasentence relation extraction can benefit from tree kernels, but intersentence relation extraction requires modeling multiple sentences or even a whole document. Incorporating special focus to entities and their related words can further improve the performance.

References
Unsupervised Parameter Estimation for OCSVMs

[Extended Abstract]

Zahra Ghafoori
The University of Melbourne
zghafoori@student.unimelb.edu.edu

Sutharshan Rajasegarar
Deakin University
sutharshan.rajasegarar@deakin.edu.au

Sarah Erfani
The University of Melbourne
sarah.erfani@unimelb.edu.edu

Shanika Karunasekera
The University of Melbourne
karus@unimelb.edu.edu

Christopher Leckie
The University of Melbourne
caleckie@unimelb.edu.edu

ABSTRACT
We propose Quick Model Selection (QMS) and Revised Duplex Max-margin Model Selection (RDMMS) methods to analyse structure of unlabelled data in order to estimate the hyper-parameters of One-Class Support Vector Machines (OCSVMs) in an unsupervised and efficient manner.

CCS Concepts
• Computing methodologies → Machine learning; Learning settings;

Keywords
OCSVM; Outlier detection; Parameter estimation.

1. INTRODUCTION
Abnormal patterns in a data set, which are inconsistent with the majority of the data, are referred to as outliers or anomalies. The hyper-plane OCSVM [2] and hyper-spherical SVDD [3] algorithms are the two widely used unsupervised outlier detection methods if there are low fractions of such patterns, otherwise they produce models that are skewed towards outliers. Availability of a labelled data set to perform cross-validation and grid-search over parameter space can fix this, but has high memory and time complexity and voids the main aim of the OCSVMs: "unsupervised outlier detection". On the other hand, a few heuristic approaches have been proposed to set hyper-parameters of OCSVMs in an unsupervised manner. This paper addresses this problem by proposing two unsupervised and efficient hyper-parameter estimation methods.

2. PURPOSED APPROACH
OCSVMs use a kernel function, such as RBF, to transform the data to a new feature space wherein outliers can be separated from normal points using a hyper-plane or a hyper-sphere (Figure 1).

Hyper-plane OCSVM:
\[
\min_{a^T K a} \; a^T a = 1; \\
0 \leq a_i \leq 1/(\nu) \forall i.
\]

Hyper-spherical SVDD:
\[
\min_{a^T K a - a^T \text{diag}(a)} \; a^T a = 1 \quad \text{and} \quad 0 \leq a_i \leq C \forall i.
\]

There are two unknown data-dependent hyper-parameters that should be adjusted: the rejection rate \(\nu\) that reflects the fraction of outliers in the training set, and the kernel parameter \(\gamma\), which smooths the learned density of the normal data. In general, our proposed methods use the following heuristic (Figure 2): (1) Computing the average distance of each data instance to its \(K\)-NNs, called \(s_K\), (2) Sorting the result vector as \(S_K\), (3) Detecting the knee-point in \(S_K\) that shows a sudden change in the data density, and (4) Setting the value of the \(\nu\) and \(\gamma\) parameter based on position and density of the knee-point. Further details on the proposed methods can be found in [1].

3. EXPERIMENTAL EVALUATION
We evaluate our proposed methods in comparison with supervised and unsupervised OCSVM parameter estimation methods on both real and synthetic data sets. The results show that our methods outperform existing unsupervised methods in terms of accuracy and time-complexity. Moreover, QMS has comparable accuracy to supervised method, while it is in average more than 900 times faster. For more detail on the evaluation results, please refer to [1].

4. REFERENCES
Reproducibility of Genomic Workflows

[Extended Abstract]

Sehrish Kanwal, Andrew Lonie, Richard Sinnott, Farah Zaib Khan

kanwals, alonie, rsinnott, khanf1@unimelb.edu.au

University of Melbourne

ABSTRACT

Reproducibility is one of the core principles for any scientific workflow yet a challenge which is not fully addressed. In this paper, a complex variant calling Genome Analysis Toolkit (GATK) workflow is implemented to investigate reproducibility requirements across three exemplar approaches to workflow definition and execution, namely programmatic, semi declarative and fully declarative.

CCS Concepts

Applied Computing - Life and medical sciences (Bioinformatics)

Keywords

Reproducibility; Workflow; GATK; Programmatic; Declarative

1. INTRODUCTION

Several Workflow Management Systems (WMSs) are available to support implementation and execution of genomic workflows. Reproducibility is one of the core principles for any scientific workflow yet a challenge which is not fully addressed. This is due to incomplete understanding of reproducibility requirements and assumptions followed by WMSs.

2. METHODS

In this research, we have implemented a variant calling workflow using three different workflow environments: Galaxy [1], Cpipe [2] and CWL [3]. In essence, the three systems represent broad range of frameworks on the WMS spectrum. The issues related to incomplete declaration and the assumptions followed in the Cpipe and Galaxy workflow were dealt with by implementing a similar workflow using CWL.

3. RESULTS

As a result of declarative implementation of GATK workflow using CWL, we identified additional set of information required to fully declare a workflow. All these details are vital to reproducibility of computational genomic analysis and hence should be documented to ensure complete transfer of information to the next user. One possible way of documenting these important bits of information is through pictorial representation of the workflow as indicated in the Figure 1.

4. DISCUSSION

Reproducibility of scientific research is increasingly important for the scientific community, as validation of scientific claims is a first step for any translational effort. Towards these objectives, this work has investigated representative workflow management systems from each end of the spectrum to propose a workflow skeleton encompassing declarations necessary to implement requirements for reproducibility. Reproducibility of an experiment often requires replication of the precise software environment including the operating system, the base software dependencies and configuration settings, under which the original analysis was conducted. Hence, the workflow implementation and reuse requires nontrivial efforts depending on the management system selected for an analysis.

5. CONCLUSION

Our case study depicts the variability in workflow implementation depending on the platform selected and aids in discovery of crucial requirements for reproducibility, missing from final workflow in case of programmatic and semi declarative approaches. It also emphasizes on the use of workflow specifications that support declaration of these reproducibility requirements. Through the implementation of GATK workflow in CWL, we have identified artefacts crucial to reproducibility to generate a complete pictorial representation of the GATK workflow. We propose provision of such pictorial representation as part of any workflow specification enabling user to declare each and every requirement using language that supports fully declarative implementation of the analysis.

6. REFERENCES


Figure 1: Pictorial representation of the GATK workflow
Improving electricity consumption prediction in a campus environment

[Extended Abstract]

Pasan Karunaratne
University of Melbourne
pkarunaratne@student.unimelb.edu.au

Shanika Karunasekera
University of Melbourne
karus@unimelb.edu.au

Aaron Harwood
University of Melbourne
aharwood@unimelb.edu.au

Masud Moshtaghi
University of Melbourne
masud.moshtaghi@unimelb.edu.au

ABSTRACT

In this extended abstract, we discuss the problem of day-ahead forecasting of electricity load consumption in a campus environment and the unique challenges it entails.

CCS Concepts

• Mathematics of computing → Time series analysis;
• Information systems → Clustering;
• Hardware → Smart grid;

Keywords

Electricity consumption prediction, Clustering, Smart Meters, Forecasting

1. INTRODUCTION

Smart meter technology has seen increased adoption in recent years, especially as a part of smart grid initiatives. Many buildings are being fitted with Smart Meters which provide information on electricity usage at fine granularities of 15 minutes or even less. This vast volume of electricity consumption data becoming readily available has led to a consequent increase in research interest in analysing Smart Meter data, especially given its potential to have direct real-world impact.

Most of this interest has centered around the analysis of data collected from consumer Smart Grids. However, it is seen that considerably less attention is placed on Smart Meter data collected in an entirely different setting to consumer use - which is in a campus of buildings. For instance, this may be a collection of buildings in a large organisation or the buildings of a university.

The electricity load data of a campus environment would differ from that of consumer data, in that the usage patterns would vary significantly from building to building depending on the particular use of the building. For example, the usage pattern of a university gymnasium would be rather different from that of a university lecture theatre. This increased variability will also be seen within the usage patterns of a single building across multiple days as well, since building utilisation will vary heavily based on the events of the day. This is in contrast to consumer power usage, which would conform to the daily and weekly rhythms of the inhabitants of the household.

Therefore analysis and forecasting using data from a campus environment presents new challenges that need to be addressed. Further, dedicated facility managers for campus buildings are more likely to translate insights into changes with direct impact on electricity consumption. Hence the analysis of campus building data presents unique opportunities as well.

2. EXPERIMENTS AND RESULTS

We analyse the data obtained from 30 smart meters installed in the University of Melbourne Parkville campus. We have observed, for the same forecasting technique [1], that the variability of the campus electricity consumption results in significantly inaccurate next-day load forecasts as compared to the forecasts for consumer consumption data obtained from the Australian Energy Market Operator. We aim to form clusters of buildings, and use the consumption information of buildings in the same cluster in order to aid the forecasting task, especially in the case of missing historical load data. We have observed that the clustering tendency using the raw data representations of the buildings is low, and therefore we are also investigating alternate representations which improve clustering tendency.

3. REFERENCES


Figure 1: Comparison of forecast errors
Investigating the Behaviour of Word Vector Models on Smaller Datasets

[Extended Abstract]

Sameendra Samarawickrama
The University of Melbourne
ssamarawickr@student.unimelb.edu.au

Shanika Karunasekera
The University of Melbourne
karus@unimelb.edu.au

Aaron Harwood
The University of Melbourne
aharwood@unimelb.edu.au

Rao Kotagiri
The University of Melbourne
kotagiri@unimelb.edu.au

ABSTRACT

Word representation models such as word2vec, have been mainly applied on large datasets and have shown to give meaningful results. While models seem to stabilize easily when trained on large datasets, we have experienced unstable results when models are trained on smaller datasets. Thus, in this paper we study this behaviour of word2vec on smaller datasets.

CCS Concepts

• Information systems → Data mining; Retrieval models and ranking;

Keywords

Word2Vec; Twitter; Social Media; Word Embeddings

1. INTRODUCTION

Research on word representation models has gained much attention lately. Word2vec [2], one such widely cited model represents each word as a vector in n real dimensions (word embeddings). Directions of these word vectors tell us how close the words are in their syntactical or semantic meaning. The literature is rich with many references to the usage of word2vec but it has been mostly used on larger datasets. For example, [2] trained with a dataset containing 1.6 billion words and [1] trained models with 300 million and 2.2 billion words. It is said that with a big enough dataset, model will eventually stabilize after several iterations as the relative positions of the input and output vectors eventually stabilize. Our initial experimental results showed that results are very unstable when trained on small datasets; giving us different results, even if the model is trained with same parameter settings. Motivated by this, in this paper we study how word2vec behaves with smaller training data and how change of parameters can lead to stable models.

Approach: First we train word2vec on a small dataset of tweets and find the most similar words to a given set of query terms (Q), using the cosine similarity. This is repeated for n number of times which resulted in n models trained with the same configuration. Then the number of common words for each query in Q, across the n models is used as an indication of the stability of the model: more common words indicates high stability. This whole process is then repeated by varying different parameters to find which configuration leads to more stable models.

2. EXPERIMENTS AND RESULTS

Our dataset contained 1789 tweets and 11714 tokens after pre-processing. We have used skip-gram model and negative sampling, as this combination have shown to give better results [2]. For each parameter setting, we trained 10 models and for each model we got 100 most similar words to 20 query terms. Figure 1 shows the histograms of word frequencies for 4 such parameter settings. All the models were trained with 5000 iterations while changing the size of the word vectors. Our results show that models become more stable as you train them with smaller vector sizes.

3. REFERENCES


Machine-Learning Algorithms Predict Graft Failure Following Liver Transplantation

[Extended Abstract]

Yamuna Kankanige*
Department of Computing and Information Systems
University of Melbourne
ykankanige@student.unimelb.edu.au

Lawrence Lau*
Department of Surgery
Austin Health

Benjamin Rubinstein
Department of Computing and Information Systems
University of Melbourne

Robert Jones
Department of Surgery
Austin Health

Christopher Christophi
Department of Surgery
Austin Health

James Bailey
Department of Computing and Information Systems
University of Melbourne

ABSTRACT

Ability to predict graft failure or primary non function at liver transplant decision time allows utilization of scarce resource of donor livers, while ensuring that patients who are urgently in need of a liver transplant are prioritized. An algorithm that is derived to predict graft failure using donor as well as recipient factors, based on local datasets, will be more beneficial in the Australian context. In this paper, we propose that machine learning algorithms can be used to achieve that objective.

CCS Concepts

• Applied computing → Health care information systems; Health informatics;

Keywords

Predicting liver transplant outcomes, machine learning, random forest, artificial neural networks

1. BACKGROUND

The decision of transplanting the liver of a donor to a recipient is a critical decision that should be made in an informed manner, not only because these decisions are affecting critically ill patients, but also because of the financial and regulatory pressures the hospitals face. Therefore, being able to predict a probability of success of a transplant, would be valuable for recipients as well as surgeons.

Scoring indices such as the donor risk index (DRI), model for end stage liver disease factors (MELD) score and survival outcomes following liver transplantation (SOFT) score have been proposed for this purpose with conflicting results[1].

2. METHODS

Liver transplant data from the Austin Hospital, Heidelberg, Melbourne, Australia, from 2010 to 2013 had been included in the study, to predict graft failure within 30 days. After considering all available donor, recipient and transplant factors (276 in total), the top 15 factors that influence the outcome were selected using a random forest based feature ranking method. DRI, MELD score and SOFT score had been used as comparative predictors of outcome.

3. RESULTS

A comparison of all the results with the study dataset is given in figure 1.

Figure 1: ROC curve comparison of different models created during the study

4. CONCLUSIONS

Using donor, transplant and recipient features which are known at the decision time of a transplant, high accuracy in matching donors and recipients can be achieved, potentially providing improved graft survival outcomes.

5. ADDITIONAL AUTHORS

Additional authors: Vijayaragavan Muralidharan, Department of Surgery, Austin Health

6. REFERENCES

ABSTRACT

Privacy disclosure has been recognized as a big challenge to health record linkage. Techniques such as k-anonymity have been proposed to mitigate the risk however they are often challenged within distributed environments due to the semantic ambiguity and implicit knowledge of heterogeneous security policies. To tackle this, we propose an improved linkage solution by leveraging semantic web technologies where privacy leakage can be detected through semantic reasoning.

CCS Concepts

- Security and privacy ➔ Access control; Privacy-preserving protocols; • Theory of computation ➔ Semantics and reasoning;

Keywords

Semantic web technologies; Privacy; Record linkage;

1. INTRODUCTION

Records linkage is a technique that allows researchers to access and use health data at remote sites whilst protecting confidential information. Through approvals of data custodians and ethics committees, researchers can access linked records. One example of such as solution is from the Centre for Health Record Linkage (CHeReL) [1]. However, with a growth of data, it is increasingly challenging to provide privacy-preserving environments where de-identification can be guaranteed from potential re-identification by linking seemingly anonymised data with other data resources. Towards this issue, we propose an automatic solution and show how semantic web technologies can be leveraged to detect privacy leakage from linkage results.

2. PROBLEM STATEMENT

Existing solutions such as k-anonymity and its variants reduce re-identification risks. Specifically, cells of key columns (quasi-identifiers) are adjusted in order to maintain the size of equivalent classes lower than a threshold. However, combining query results from discrete repositories might cause the loss of confidentiality. For instance, patient data potentially includes other information such as the country of birth, ethnicity, postcode etc. whilst seemingly safe, such information can be used for re-identification, e.g. where countries of birth and ethnicities seldom occur in particular postcodes. Ideally, such privacy issues should be identified and checked before policies are violated.

3. SOLUTION TO PRIVACY LEAKAGE

Linked open data is a way of publishing data/metadata enriched with semantic associations, so that web content can be searched in different representations between related resources [2]. Similarly, knowledge released by authorities can be comprehensively used in risk assessment. Based on the privacy protocol of CHeReL, Figure 1 shows a semantically enabled architecture. Prior to exchanging information, staff at Centre B collect metadata from participant organizations (A & C) and translate it into a linked format (e.g. .rdf). Based on the request from A, queries are redirected and identifying information pulled from related custodians (like C). Instead of making local decisions, custodians can delegate the overall risk analysis to trusted centers by submitting aspects of their privacy policies. One common data disclosure risk is geospatial re-identification. In this work we show how semantic technologies allow supporting such reasoning at geospatial aggregation levels in an intelligent manner.

Figure 1. Overview of semantic disclosure control model

Driven by projects such as the Australian Urban Research Infrastructure Network (AURIN) and the Australian Type-1 Diabetes Data Network (ADDN), we consider a scenario where densities within given geospatial areas are the disclosure criteria, i.e. densities of individuals in particular sub-regions such as Statistical Areas (SA4, SA3, SA2, SA1) or mesh blocks should be restricted. Such geographical vocabularies used by ADDN and AURIN can be modelled in a linked form through using semantic properties such as contains and belongsTo. On this basis, semantic reasoning can be launched on an overlaps relation where privacy issues arise. For instance, the SA1 Census for languages spoken at home indicates the numbers of different speakers at the SA1 region. Since certain languages can imply countries and ethnicities it may be the case to relate those concepts with equalTo. Often if there is a sparse distribution of certain speakers in a given SA1 region then ethnical details are withheld. Rather than simply refuse to release (link) data at such regions, however, it is possible that data could be released at a higher level (SA2). Since SA1 regions are included within SA2 regions, it is possible to automatically aggregate the linkage at geospatial classifications that meet the density criteria. Therefore, instead of simply releasing potentially re-identifying information, it is possible to reason on the linked knowledge to support richer security scenarios. In this case, based on the census counts, speaker densities can be calculated at the SA1 level. For linkage regions having a higher density than the related areas of interest, certain information needs to be generalized/suppressed (e.g. Chinese -> North East Asian) in order to protect data subjects from potential re-identification.

4. REFERENCES

[1] Centre for Health Record Linkage (CHeReL)

A segmentation method for counting repeats in a weightlifting routine

[Extended Abstract]

Yousef Kowsar
kowsar.yousef@student.unimelb.edu.au

Eduardo Velloso
eduardo.velloso@unimelb.edu.au

Lars Kulik
lkulik@unimelb.edu.au

Christopher Andrew
Leckie
caleckie@unimelb.edu.au

Masud Moshtaghi
masud.moshtaghi@unimelb.edu.au

Department of Computing and Information Systems
The University of Melbourne

ABSTRACT

Weightlifting injuries often occur as a result of deviating from the correct form of an exercise. Thus, designing a system that can alert its users, when deviating from a safe margin during their exercise routine has been the focus of recent studies. The first step for such a system is to identify what exercise the user is performing. Tangibility of motion sensors through most sport gadgets, where users can easily purchase an off the shelf tool makes them an appealing solution for alerting users. In this project, we aimed to use the data collected through a sport watch worn on the user’s forearm to detect the exercise. We showed how segmentation can help improving the exercise identification task by validating the result through a real world dataset.

CCS Concepts

• Human-centered computing → Ubiquitous and mobile computing design and evaluation methods;

Keywords

Motion sensor networks, Weightlifting

1. METHOD AND RESULT

Recent advances in sensor hardware make them ubiquitous in daily life. Sensors such as accelerometer have become standard feature on most sport gadgets. As a result, their usage for tracking gym exercise activities has been under many studies. Cheng et al. [1] showed that motion sensors will provide a valuable feedback on both aspect of exercise classification and quality of performance. Hence, the main challenge to build a system that can alert the user, while performing an exercise, is to differentiate between the correct form of a posture from the incorrect form. Specially, by considering the fact that weightlifting exercises are repetitive tasks, where an exercise will repeat for a certain number of cycles, the user might perform an exercise correctly for a few repeats and then start deviating from the correct form. In this regards, Velloso et al. [2] introduced the idea of classifying exercises by error type. They defined a set of postures as the most common errors people will do during a specific exercise. Through an empirical study, they showed that using on-body motion sensors, it is possible to classify the error type performed by the user, using a random forest classifier. For their study, they defined 4 common errors of unilateral curl bicep, labeled Move B to Move E respectively. They also collected the data for the correct form of the unilateral bicep Move A. They collected the data through 6 different people performing each move for 10 repeats. The collected data was from 4 sets of accelerometer, gyroscope and magnet sensors placed on dumbbells, forearm, upper arm and belt of each user. They segmented the sensor data with a fixed length window and classified every single point in each segment to one of the five different labels. Their method has two main problems. First, a segment should only have one label assigned to it. Segmenting each data read might result in mis-representing the classification task. Second, the fixed window technique will lose the track of each repetition by mixing the repeats performed by the user. In this paper, we derived a mathematical model based on accelerometer’s function derivatives to segment the data correctly. We used a random forest classifier to classify each segment. We showed that it is possible to receive 99.9 percent precision for most classes, while in the worst case scenario we got 90 percent precision. We showed how segmenting the temporal data using only forearm accelerometer will improve the classification task for each class. Our result showed that for the analyzed dataset, accelerometer data will outperform the current technique for classifying the task, which in turn reduces the required number of sensors from 12 to a single accelerometer placed on the forearm.

2. REFERENCES


Dear Noise: Hide Me Please!

[Extended Abstract]

Elham Naghizade, James Bailey, Lars Kulik, Egemen Tanin
Computing and Information Systems Department
The University of Melbourne
Melbourne, Australia
{enaghi,baileyj,lkulik,etanin}@unimelb.edu.au

ABSTRACT
Differential privacy has been deployed and adapted in different application scenarios due to its rigorous protection of individuals’ privacy regardless of the adversary’s background knowledge. Most of the proposed solutions in this area focus on releasing private counters or histograms, which involve low sensitivity, and the main focus of these solutions is minimizing the amount of noise and the utility loss throughout the process. In this paper we consider the case of releasing private numerical values with unbounded sensitivity in a dataset that grows over time. We show that straightforward application of current mechanisms cannot guarantee (differential) privacy for individuals under an open-world assumption, especially if the dataset is updated by an outlier.

CCS Concepts

•Security and privacy → Domain-specific security and privacy architectures;

Keywords
Differential Privacy; Unbounded Global Sensitivity; Dynamic Datasets

1. INTRODUCTION
Ubiquitous sensing technologies and novel analytical solutions have paved the way for applications from health tracking to customised location-based services. Quality of data is key to the data analysis, however, exploiting precise data may lead to serious breaches in privacy. This has encouraged a range of studies on how to balance privacy against the quality of analysis, i.e., data utility. Differential Privacy (DP) [3], that aims to preserve users’ privacy through guaranteeing that the presence or absence of a user in the dataset does not have a considerable effect on the outcome of a query. The presence or absence of a user determines the global sensitivity, which intuitively models the maximum difference a single user can make wrt a certain query.

However, most existing approaches that provide differential privacy focus on a static dataset, i.e., they focus on a closed-world scenario. This makes DP solutions impractical for people-centric scenarios like participatory sensing networks where people constantly leave and join the dataset, hence, rapidly changing the dataset. The study of differentially private releases of time-series or updated outputs has recently gained attention. Authors in [2, 4] propose differentially private counter releases under continual observations.

The study in [1] proposes a method that handles dynamic leaves and joins, e.g., nodes in a sensing network where each node encrypts and perturbs its value and a trusted aggregator computes the noisy sum of the values. [5] develops a mechanism to release differentially private histograms in a dynamic setting.

Nonetheless, the current state-of-the-art focus on cases with bounded and small global sensitivity, e.g., counts of events at a certain time. We focus on the provided privacy levels for the case of differentially private release of numerical values (with unbounded sensitivity), e.g., average salary, average daily power consumption of a household, etc. The continual release of this information is of great importance in many areas, e.g., urban planning applications.

In our work, we show that it is not possible to simply observe a changed average value for a successful attack against a differentially but evolving private release: The effect of adding an outlier to the dataset is smoothed out by a growing dataset. Further, adding noise proportional to the size of the population ensures small deviations from the true average results for large datasets. Consequently, an adversary is not able to detect an outlier simply from the fluctuations in the average results. We then provide a successful attack that uses prior private query results as background knowledge of an adversary to gain estimates of the underlying noise distribution of each release and determine whether or not an outlier has joined the dataset.

In summary, we provide both formal and empirical evidence that although data dependent noise maintains utility, which makes the data exchange process meaningful/possible, it cannot guarantee the promised level of privacy if new data is added to an existing dataset. This calls for a rethink of current DP mechanisms that deal with numerical values of users to ensure user privacy in a dynamic world.

2. REFERENCES
Supervised Learning Meets Biocuration: Detecting Duplicate Records in Biological Databases

[Extended Abstract]

Qingyu Chen
Department of Computing and Information Systems
Parkville VIC 3010
Melbourne, Australia
qingyuc1@student.unimelb.edu.au

Justin Zobel
Department of Computing and Information Systems
Parkville VIC 3010
Melbourne, Australia
jzobel@unimelb.edu.au

Karin Verspoor
Department of Computing and Information Systems
Parkville VIC 3010
Melbourne, Australia
karin.verspoor@unimelb.edu.au

ABSTRACT
We develop a duplicate detection method for biological databases using supervised learning. Our experiments show that supervised learning has substantial promise in biocuration.

CCS Concepts
- Applied computing → Bioinformatics;
- Information systems → Relational database model;
- Computing methodologies → Classification and regression trees;

Keywords
Duplicate detection; Bioinformatics; Data curation

1. INTRODUCTION
A duplicate in databases is an instance where more than one record represents the same entity. Detecting duplicate records is a key and often the first step in biological databases. In general de-duplication in databases have two kinds: (1) Accuracy, detecting duplicates in an accurate manner; (2) Efficiency, detecting duplicates in an efficient manner. In highly curated databases like Swiss-Prot, accuracy is the most important priority [1]. Supervised learning has the potential to detect duplicates precisely by building automatic systems that learn from expert curation. While machine learning is a mature approach in other duplicate detection contexts, it has seen only preliminary application in biological databases.

We developed and evaluated a supervised duplicate detection method based on an expert curated dataset of duplicates, containing over one million pairs across five organisms derived from genomic sequence databases. We selected 22 features to represent distinct attributes of the database records, and developed a binary model and a multi-class model. Both models achieve promising performance; under cross-validation, the binary model had over 90% accuracy in each of the five organisms, while the multi-class model maintains high accuracy and is more robust in generalisation experiment.

2. METHOD
We collected duplicate and distinct record pairs across five organisms labelled by Swiss-Prot experts. For each pair, we extracted 22 features from each pair in three categories: (1) meta data, annotations of the records; (2) sequence, sequences of records; (3) coding regions: coding sequences and translations of records. Then we employed Naive Bayes, Decision Tree and SVM to develop both binary-class and multi-class model, where in the latter model duplicates are categories into different types. To further evaluate the method, we examined the learning curve to see how much data is sufficient for training; an ablation study to quantify the impacts of different features; error analysis to investigate what cases are difficult to detect automatically; generalisation to find out whether the trained models is applicable in broader context. An example of learning curve measurement on SVM is illustrated at Figure 1.

Both models have over 90% accuracy in all five organisms. The learning curve measurement shows the method has the potential to be more efficient by only training sample of the dataset. We also found that features for meta-data, sequence similarity, and quality checks on alignments achieved the best results. The multi-class models are more robust in the generalisation experiment.

3. REFERENCES

Figure 1: Evaluate the learning curve of SVM on Danio rerio
Automatic Detection of Choroid-Sclera Interface in EDI-OCT Images

[Extended Abstract]

Md Akter Hussain
The University of Melbourne
hussain@student.unimelb.edu.au

Alauddin Bhuiyan
New York University
bhuiya02@nyu.edu

Kotagiri Ramamohanrao
The University of Melbourne
kotagiri@unimelb.edu.au

ABSTRACT
The choroid, located next to retina, plays an important role in the pathology of retinal diseases. We propose a method for segmenting the outer choroid boundary called choroid sclera interface (CSI) from Enhanced Depth Imaging Optical coherence Tomography (EDI-OCT) images.

CCS Concepts
• Computing methodologies → Image processing
• Computing methodologies → Image representations

Keywords
Image Processing, EDI-OCT, Shortest path, Choroid, Retina.

1. INTRODUCTION
Optical coherence tomography (OCT) is an emerging non-contact and non-invasive imaging method. OCT enables a high resolution, three-dimensional (3D) cross-sectional imaging of the retina and choroid (vascular layer connecting retina and sclera) [1]. The quantitative and qualitative evaluation of the choroid is important in order to understand its relationship with the optic nerve and retinal diseases such as glaucoma, and age-related macular degeneration (AMD) [2]. We propose an automatic method for segmenting choroid-sclera interface (CSI) from EDI-OCT images and adapted a method from [3] for segmenting the inner boundary of the choroid called the retinal pigment epithelium, Bruch’s Membrane and Choriocapillaris (RBC) complex boundary.

2. PROPOSED METHOD
The algorithm is based on graph shortest path for segmenting and detecting the CSI. For increasing the efficiency and accuracy of the detection, a small region of interest (ROI) is selected first automatically. Depth based intensity normalization is applied for equalizing the intensity of the surface pixels of the choroid and sclera. The pixels of the image form the node of the graph and shortest path algorithm is used to find the CSI.

3. Evaluation
We used 190 B-scan (2-D) images of 10 subjects for evaluating our method against manual grading. Fig. 1 shows segmentation of our method. Table 1 reports the root mean square error for the RBC and CSI as well as the dice coefficient of the choroid region.

4. CONCLUSION
We proposed a robust, highly accurate automatic method for the detection of CSI. Our method has the level of performance of an experienced grader in terms of accuracy and the inter grader variability. Performance evaluation indicates our automatic method can be applied for large scale study of retinal disease detection, progression and diagnosis related to choroid.

5. REFERENCES
The Hitchhiker’s Guide to the Optimal Route Planning

Oleksii Vedernikov
Supervisors: Lars Kulik, Kotagiri Ramamohanarao
Department of Computing and Information Systems
The University of Melbourne
o.vedernikov@student.unimelb.edu.au

ABSTRACT
This paper analyses hitchhiking from a computational perspective. We formulate an optimization problem of finding the path with the lowest expected time for a hitchhiker, and provide a robust algorithm to solve it.

CCS Concepts
• Theory of computation → Shortest paths;

Keywords
Hitchhiking, shortest path, dynamic programming

1. INTRODUCTION
Hitchhiking is a popular tourism approach widely spread all around the world. It represents getting rides on random cars towards the desired destination. From a mobility point of view, hitchhiking differs from travel by private vehicles or public transport due to high uncertainty of both route and travel time. While riding time for hitchhikers is identical to that of drivers, waiting time to catch the cars varies significantly according to locations as well as the itineraries of drivers passing by. While non-optimal route planning may cause delays for car drivers, the same for hitchhikers may result in hours of hopeless waiting at empty roads and induce safety risks.

Despite its importance and prominence, to the best of our knowledge, the idea of efficient route planning for hitchhikers has never been studied. Consequently, the main research question of this work is: on a road network, how to find an optimal hitchhiking route between two cities?

2. PROBLEM FORMULATION
Even though hitchhiking trips may vary in terms of safety, cost, comfort etc, we assume the uniformity of these parameters and narrow the focus to the most important: total travel time. We aim to find not only an optimal path, namely a sequence of roads, but also destinations of cars that they should pick up on their way.

Therefore, we need to transform a road network to a hitchhiking graph: $HG = (V, E, W, R)$, where $V$ is a set of nodes, $E$ is a set of edges, and $W = \{w(v_i, v_j, e_k) | \forall v_i$ can reach $v_j; e_k$ is incident to $v_i\}$ given in Figure 1.

Despite its importance and prominence, to the best of our knowledge, the idea of efficient route planning for hitchhikers has never been studied. Consequently, the main research question of this work is: on a road network, how to find an optimal hitchhiking route between two cities?

3. CONCLUSION
The key contributions of this paper are creating a novel concept of a hitchhiking graph in conjunction with an efficient algorithm for finding the optimal route with the least expected time between any two points at a given start time. The robustness of the algorithm is studied on the real road networks and synthetic graphs.
An Efficient Algorithm for Non-Interactive Differential Privacy

[Extended Abstract]

Soheila Ghane, Lars Kulik, Ramamohanarao Kotagiri
Department of Computing and Information Science
The University of Melbourne, Australia
{soheila.ghane,lkulik,kotagiri}@unimelb.edu.au

ABSTRACT
Differential Privacy is one of the state of the art privacy models for publishing data which formally guarantees individual privacy. However Differential Privacy does not promise the released data will have sufficient utility for further analysis. To publish data with privacy and utility, we propose a mechanism based on Differential Privacy which achieves near optimal theoretical privacy bounds. Experiment demonstrate our mechanism performs efficiently and accurately in comparison to existing approaches.

CCS Concepts
• Theory of computation → Pattern matching;

Keywords
Differential privacy; Publishing Data, Utility

1. OVERVIEW
In releasing data, preserving privacy of sensitive data is an important objective. Promising no private information leakage and guaranteeing usefulness for data analysis are two opposing criteria. Recently, differential privacy has emerged which provides a formal and provable definition of privacy. However, techniques developed to achieve this privacy definition do not guarantee usefulness of released data and the techniques are not efficient. The key challenge is proposing a mechanism that preserves privacy of sensitive data and guarantees the output has adequate utility for data analysis.

We propose an efficient mechanism which guarantees a lower bound on utility of released data. This algorithm approximates distribution of original data and improves it with respect to and query errors. Subject of this algorithm is the class of linear queries, which return a weighted sum of the elements in the data. In each iteration the worst query identifies how much the correction should be made and then values in the realised data are modified using the multiplicative weights update rule. In terms of privacy guarantee and utility bound, this algorithm substantially improves over previous works[1]. Figure 1 displays how the proposed algorithm outperforms a recent work, MWEM, [2] on 10,000 sample dataset with 7 binary attributes. Also, it runs faster than MWEM. Let Q be the set of queries q, B be the original data and A be approximated distribution, KL-Divergence measures the difference between true distribution B and approximated distribution A:

\[ KL - Div(B||A) = \sum_i B(i) \log \frac{B(i)}{A(i)} \]

Accuracy and utility of results are computed as follows:

\[ Utility = 1 - \frac{1}{|Q|} \sum_i \left| \frac{q_i(B) - q_i(A)}{|q_i(B)| + |q_i(A)|} \right| \]

and

\[ Accuracy = 1 - \frac{1}{|Q|} \sum_i \left| \frac{q_i(B) + Lap(\frac{|Q|+1}{\epsilon}) - q_i(A)}{|q_i(B) + Lap(\frac{|Q|+1}{\epsilon})| + |q_i(A)|} \right| . \]

2. REFERENCES

Figure 1: Performance of proposed algorithm vs. MWEM. x-axis shows number of iterations and y-axis measures Accuracy/Utility.
Biographical Fact Extraction using non-linear SVM kernels

Gitansh Khirbat  
CIS, University of Melbourne  
gkhirbat@student.unimelb.edu.au

Rui Zhang  
CIS, University of Melbourne  
rui.zhang@unimelb.edu.au

Jianzhong Qi  
CIS, University of Melbourne  
jianzhong.qi@unimelb.edu.au

ABSTRACT
Relation extraction systems saw a rapid growth in the last decade, with the focus shifting from extracting domain specific binary relations to finding generalized relations using open Information Extraction (IE) systems. Question answering and summarizing systems have given rise to n-ary relation extraction where \( n \geq 2 \). We propose an IE system which extracts entities and n-ary relations from biographical documents through a two stage process involving entity extraction using Conditional Random Fields (CRF) and n-ary relation extraction using Support Vector Machines (SVM) with non-linear kernels. We investigate the shortest path dependency information along with other syntactic and semantic features, and report a significant improvement over the state-of-the-art system.

CCS Concepts
- Information systems → Information extraction;
- Theory of computation → Support vector machines;

Keywords
Information Extraction; Support Vector Machines; Radial Basis Kernel

1. INTRODUCTION
The goal of n-ary relation extraction systems is to find relations between more than two entities present in a sentence. There has been some work in this field where maximal cliques are used on entity-graph structures [1] and lexical semantics are used to train a distant-supervision based model. In this work, we extract n-ary relation tuples \( (n \in [2, 4]) \) from a corpus of biographical documents. Table 1 lists some of the desired relations over a set of entities like Person, Degree, University, Discipline, Organization, Position and Department. We make use of CRF and SVM-based models for entity extraction and relation extraction respectively.

<table>
<thead>
<tr>
<th>n-arity</th>
<th>Relation Tuple</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>(Person, Degree); (Person, University)</td>
</tr>
<tr>
<td></td>
<td>(Person, Position); (Person, Organization)</td>
</tr>
<tr>
<td>3</td>
<td>(Person, Degree, University)</td>
</tr>
<tr>
<td></td>
<td>(Person, Position, Organization)</td>
</tr>
<tr>
<td>4</td>
<td>(Person, Degree, Discipline, University)</td>
</tr>
<tr>
<td></td>
<td>(Person, Position, Organization, Department)</td>
</tr>
</tbody>
</table>

Table 2: Feature-wise performance of n-ary relations

<table>
<thead>
<tr>
<th>Feature</th>
<th>Precision</th>
<th>Recall</th>
<th>F1-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entity Presence</td>
<td>59.83</td>
<td>62.33</td>
<td>61.05</td>
</tr>
<tr>
<td>+ POS-tag sequence</td>
<td>64.29</td>
<td>63.67</td>
<td>63.97</td>
</tr>
<tr>
<td>+ Gazetteer</td>
<td>71.01</td>
<td>69.13</td>
<td>70.05</td>
</tr>
<tr>
<td>+ Dependency Parse</td>
<td><strong>79.57</strong></td>
<td>75.51</td>
<td><strong>77.49</strong></td>
</tr>
<tr>
<td>UPenn System</td>
<td>63.03</td>
<td>77.26</td>
<td>69.42</td>
</tr>
</tbody>
</table>

2. METHODOLOGY

2.1 Dataset
Our dataset consists of 400 biographical documents retrieved from homepages of Australian researchers. It has more than 1,300 binary relations, 1186 3-ary relations and 647 4-ary relations, along with more than 4,800 entities.

2.2 System Overview
Document pre-processing is done using the Stanford Core NLP. IE occurs in 2 stages. In the first stage, CRF-model is trained to classify words into 7 classes corresponding to the seven entities. Surface tokens, Part of Speech (POS) tags and gazetteers are used as features for training the model. The second stage is relation extraction using SVM with Radial Basis Function kernel to classify groups of entities within a sentence into relations as mentioned in Table 1. For example: “[Peter Person] received a [PhD Degree] from [Unimelb University]” gives us a 3-ary relation. Features used in this stage are the extracted entities, POS tag sequences, gazetteers and full parse tree along with shortest path dependency-based rules between entities and the connecting verb. We used 10-fold cross validation to prevent overfitting, and adopted a grid search to find \((C, \gamma)\) to be \((2^2, 2^{-3.5})\).

3. RESULTS
The entity extractor gives us an overall accuracy of 90% over the 7 entity categories. Impact of different features on relation extraction is shown in Table 2.

4. CONCLUSION
Shortest path dependency-based rules improve the n-ary relation extraction accuracy by 8%.

5. REFERENCES
A Bayesian Approach for Inferring the Hidden Truth from Crowdsourced Annotations

[Extended Abstract]

Yuan Li, Meng Fang, Benjamin I. P. Rubinstein, and Trevor Cohn
The University of Melbourne
yuanl4@student.unimelb.edu.au
{meng.fang, benjamin.rubinstein, t.cohn}@unimelb.edu.au

ABSTRACT

In recent years, crowdsourcing has become a promising technique for collecting large amounts of labelled data at a low cost. However, due to low quality annotations, items are typically redundantly labelled by several different workers, with labels aggregated subsequently. Unfortunately, naive methods such as majority voting are inefficient for uncovering true annotations. In this paper, we propose a novel method for aggregating binary labels that combines a Boltzmann machine with latent confusion matrices to model both the correlation between the latent true annotations and the reliability of workers. Experimental results on two datasets demonstrate that our proposed approach can outperform both the majority voting baseline and a Gaussian process classification model. Moreover, we show further gains when integrating a proposed active learning heuristic.

CCS Concepts

• Information systems → Crowdsourcing; • Computing methodologies → Latent variable models; • Theory of computation → Gaussian processes;

Keywords

Crowdsourcing; Latent variable models; Gaussian processes.

1. METHODOLOGY

Combining the confusion matrices and the Boltzmann machine, we define the joint distribution of binary variables \((-1, +1)\) worker labels \(Y\) and true annotations \(Z\) as

\[
P(Y, Z | \theta) = \prod_{i=1}^{N} \prod_{j \in W_i} P(y_{ij} | z_i, \theta_j) P(Z)
\]

where \(P(Z) \propto \exp(Z^T K Z)\) and \(P(y_{ij} | z_i, \theta_j)\) is defined by a \(2 \times 2\) confusion matrix for worker \(j\) as

\[
P(y_{ij} = 1 | z_i = r, \theta_j) = \theta_{ijr}, \quad \forall i \in D_j
\]

To avoid overfitting when there are insufficient worker labels, we add priors to \(\theta_{ijr}\). A natural choice is the Beta prior which is conjugate to Bernoulli \(P(y_{ij} | z_i, \theta_j)\).

\[
P(\theta | \alpha) = \prod_{j \in W} P(\theta_j | \alpha) \propto \prod_{j \in W} \prod_{r = \pm 1} \prod_{i \in \pm 1} \theta_{ijr}^{\alpha_r - 1}
\]

where \(\alpha\) are Beta hyperparameters.

Gibbs sampling estimates the mean of \(Z\) based on

\[
P(z_i = r | Y, Z^{-i}, \alpha) \propto \left( \prod_{j \in W_i} n_{ijr} + \alpha_{ir} \right) \left( \sum_{\forall j \neq i} k_{ijr} r z_i \right)
\]

2. EXPERIMENTAL STUDY

Table 1: Results on the full sentiment polarity dataset.

<table>
<thead>
<tr>
<th>Method</th>
<th>Trainset Accuracy</th>
<th>Testset Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GPC-GOLD†</td>
<td>0.987</td>
<td>0.723</td>
</tr>
<tr>
<td>LR-GOLD#*</td>
<td>0.822</td>
<td>0.732</td>
</tr>
<tr>
<td>GPC-MV†</td>
<td>0.886</td>
<td>0.719</td>
</tr>
<tr>
<td>GPC-MA†</td>
<td>0.900</td>
<td>0.721</td>
</tr>
<tr>
<td>MV*</td>
<td>0.8851</td>
<td>0.6634</td>
</tr>
<tr>
<td>BMMA*</td>
<td>0.9140</td>
<td>0.6950</td>
</tr>
<tr>
<td>BMMA&amp;LR*</td>
<td>0.9140†</td>
<td>0.7246†</td>
</tr>
</tbody>
</table>

# Numbers shown in bold are the best results.

5 Methods trained on true annotations instead of worker labels. Their results are in italic and provide upper bounds for others.

† Results copied from Rodrigues et al. [2, 1].

3. REFERENCES


Distributed Coding of time series data

[Extended Abstract]

Lakshmi J Mohan
Research student
ljagathamma@student.unimelb.edu.au

Dr. Udaya Parampalli
Assoc. Professor
udaya@unimelb.edu.au

Dr. Aaron Harwood
Senior Lecturer
aharwood@unimelb.edu.au

ABSTRACT
As volume of big data generated worldwide increases everyday at an unforeseen rate, novel techniques are required to make the storage and processing of such huge volumes of data more efficient. This paper discusses about one such idea in the context of video generated by sensor devices that are located at the periphery of an Internet of Things (IoT) architecture. We propose design of a binary tree based model for coding time series data. The paper explains the design of the system and related research questions.

Keywords
Distributed coding; binary tree; IoT

1. INTRODUCTION
This short paper explains the design of a binary tree based coding solution for storing distributed time series data. By distributed time series data we mean data getting generated by any device in the periphery of an IoT set-up, like a camera. These devices generate data frames at successive units of time and require huge amount of storage. The assumption taken here is that old data frames are very rarely accessed and their access speed is not of much concern. Hence, they can be coded in such a way that it will take longer to retrieve them. On the other hand, newer frames, e.g., frames generated during the past few hours, are quite likely to be accessed and hence need to be coded for enabling faster retrieval. Thus, we need a system that can adapt itself to user requirement of hot and cold data.

2. PROPOSED MODEL
The binary distributed coding system consists of a binary tree where leaf nodes generate raw frames and internal nodes generate coded frames. Let $V_j$ be the set of nodes at level $j$. The root node is level 0. Let $l_u$ be the level that node $u$ is at in the tree. Let $p_u$ be the parent of node $u$ for all nodes $u$ that are not the root. Let $c_{u,1}, c_{u,2}$ be the two children for node $u$, for all nodes but the leaf nodes. All nodes are said to generate frames (either raw or coded), where $f_{u,i}$ is the $i$-th frame for node $u$, and $i = 1, 2, \ldots$ is an unbounded frame index. Let the time that a frame is generated be $t_{u,i}$. A parent node determines for each equal frame number of its two children, at what time in the future the frame will be dropped. Let $\delta_{u,i}$ be an exponential random variable with parameter $\lambda_{u,i}$. Only one of the children drops the frame, and this is picked uniformly at random by the parent.

The drop time for frame $i$ is:

$$t_{u,i} = \max[t_{c_{u,1},i}, t_{c_{u,2},i}] + \delta_{u,i}$$

We take the maximum because children may not produce frames at exactly the same time. Notice that $t_{u,i}$ is the time that node $u$ generates its coded frame $i$. Following the above system definition, applied to all nodes at all levels (note that the root never drops frames), for any given frame (raw or coded), generated at time $t$, the probability that it has been dropped by time $t + \Delta$ is:

$$P(t+h = \Delta) = 1 - e^{-\lambda \Delta}$$

The factor $\frac{1}{2}$ represents the fact that for each pair of children, only one of them actually drops the frame. We can thereby calculate the probability distribution on height:

$$P[r = h; u, \Delta] = \prod_{t=h+1}^{l_u} \frac{1}{2} (1 - e^{-\lambda \Delta})(1 - \frac{1}{2}(1 - e^{-\lambda u-h \Delta}))$$

Here $P[r = h; u, \Delta]$ is the probability that reconstruction of a frame at node $u$ at a time $\Delta$ after the frame was constructed, requires contacting up to $h$ parents. For $h = 0$ the above equation simplifies to the probability that the frame has not been dropped at the node (i.e. no reconstruction is required). The value of $h$ can range between 0 and $l_u$. Since reconstruction will require message passing back down the tree, and assuming that message passing can occur concurrently down different sub-trees, the total delay for reconstruction is bounded by: $4h$ where $h$ is the average height up the tree that reconstruction is required for leaf nodes.

3. RESEARCH QUESTIONS
The natural question is: when a frame is required to be reconstructed at a leaf node, how much message passing is required on average? To answer this, note that if a frame $i$ has been dropped at node $u$ then $u$ requires to ask its parent to reconstruct this frame. Similarly, the parent may have dropped the required (coded) frame and will thereby require asking its parent in turn. In a real IoT set-up, the equivalent parameters of concern would be the delay and bandwidth involved in the recovery process when a frame is requested.
Platforms and Systems
Finding Data Races in Java programs

[Extended Abstract]

D. A. W. Clarke
Dept. of Computing and Info Systems
University of Melbourne
dawclarke@gmail.com

T. Miller
Dept. of Computing and Info Systems
University of Melbourne
tmiller@unimelb.edu.au

A. Mendoza
Dept. of Computing and Info Systems
University of Melbourne
mendoza@unimelb.edu.au

ABSTRACT
Data races are avoided by enclosing accesses to shared variables within critical sections. We have devised a process for detecting data races caused by misuse of critical sections. To improve the performance of multi-threaded applications, we use Abstract Execution Graphs (AEG) to evaluate novel section de-limiter patterns that are less disruptive to the performance of weak memory CPUs such as ARM and POWER.

CCS Concepts
• Software and its engineering → Software organization and properties → Software functional properties → Correctness → Synchronization

Keywords
Multi-threaded; Java; data races; weak memory models;

1. INTRODUCTION
Last year we published a static analysis for detecting data races caused by the misuse of critical sections [1]. We have since refined that process to address the challenge of path explosion. More recently, we have applied the Abstract Execution Graph (AEG) concept to investigate patterns for critical section delimiters that cause less disruption to the performance of highly optimised popular processors, such as ARM and POWER, which provide particularly weak memory models.

2. CRITICAL SECTIONS
2.1 Data Races
Data races are never benign. They are avoided by placing accesses to shared variables within critical sections. We can identify a method as a critical section by using the synchronized construct and we can, similarly, use the synchronized construct to identify a block as a critical section.

Our previously published work has shown that, provided we can identify the starts and ends of critical sections, a static analysis can find the data races caused by the misuse of critical sections.

2.2 Alternatives to synchronized
There is anecdotal and experimental evidence that the synchronized construct is expensive. Research has concluded that informal synchronization techniques are usually flawed.

Java 5 provided the java/util/concurrent package that included classes, such as AtomicInteger, that provide a CompareAndSet method. Using these features we can devise alternative cheaper patterns for use as critical section de-limiters.

2.3 Impact of weak memory models
Sequential consistency (SC) states that the results of an execution are indistinguishable from those that would be observed if the program were to be executed as written. The x86 series of processors offers a relatively strong memory model, but the ARM and POWER processors provide much weaker memory models.

3. ABSTRACT EXECUTION GRAPH
Research into C programs at Oxford [2] has shown that a lack of SC can be detected by using a "no values" abstract similar to that proposed in our earlier work. They show that this Abstract Execution Graph (AEG) concept can be used to detect a lack of SC simply by finding cycles in the graph. For example in Figure 1, the sequence (1, 2, 3, 4, 1) is a cycle. They also report the successful implementation of an algorithm to optimise the choice of and placement of memory fences to restore sequential consistency by breaking these cycles.

Table 1 - Preserved program order versus fences by CPU

<table>
<thead>
<tr>
<th></th>
<th>SC</th>
<th>x86</th>
<th>POWER</th>
</tr>
</thead>
<tbody>
<tr>
<td>poWR</td>
<td>yes</td>
<td>mFence</td>
<td>sync</td>
</tr>
<tr>
<td>poWW</td>
<td>yes</td>
<td>yes</td>
<td>sync, lwsync</td>
</tr>
<tr>
<td>poRW</td>
<td>yes</td>
<td>yes</td>
<td>sync, lwsync, dp</td>
</tr>
<tr>
<td>poRR</td>
<td>yes</td>
<td>yes</td>
<td>sync, lwsync, dp, branch, isync</td>
</tr>
</tbody>
</table>

At a machine-code level, SC can be restored by the judicious insertion of memory fences. Table 1 shows that x86 has only one, strong memory fence, while the POWER processors have a variety of other fences that cause less disruption to execution. ARM is similar to POWER. In section 3 we describe a process for designing more "hardware friendly" section de-limiter patterns.

Figure 1 - AEG for message-passing pattern
Other theoretical research into logic for reasoning about relaxed memory models [3] has suggested that performance is better sustained by separating memory fences from the acquire-read and release-write events. We have used this principle in Java to modify algorithms like Dekker's mutual exclusion algorithm to make them efficient in weak memory environments. We then applied the AEG and fence-placement techniques to validate that the solution provides the least disruption to weak memory processors. We shall to use this technique to develop a series of better architecture-specific critical section de-limiter patterns.

4. REFERENCES
Parallel Algorithms for $k$–Center Clustering

[Extended Abstract]

Jessica McClintock
University of Melbourne
jmccclintock@unimelb.edu.au

ABSTRACT

The $k$-center problem is a classic NP-hard clustering question. For contemporary massive data sets, RAM-based algorithms become impractical. Although there exist good sequential algorithms for $k$-center, they are not easily parallelizable. We observe that Gonzalez’s greedy algorithm can be efficiently parallelized; in practice, we find that two rounds are sufficient, leading to a 4-approximation. We contrast this with an existing parallel 10-approximation algorithm for $k$-center, which we parameterize to trade off runtime for approximation guarantee. Our approach is about 100 times faster than both its sequential version and the parallel sampling approach, barely compromising solution quality.

CCS Concepts

- Theory of computation → MapReduce algorithms; Facility location and clustering;

1. INTRODUCTION

Clustering problems appear in fields such as social networking and bioinformatics in which data sets can be prohibitively large. Many such problems are NP-hard, so it is common to use polynomial-time approximation algorithms that ensure solutions are within some factor of optimality. When RAM is insufficient to perform the required calculations it becomes practical to seek parallel solutions.

Definition 1. The $k$-centre problem aims to find a set of at most $k$ vertices such that the maximum distance from a vertex to its assigned centre is minimised.

We present a parallelization scheme for $k$-centre that, given sufficient capacity relative to the size of the data, can give a 4-approximation and only requires two iterations. In practice, the solutions given by our algorithm are comparable to both a sequential algorithm [2] and an alternative parallel implementation [1], while requiring fewer iterations and avoiding many of the restrictions of other approaches.

2. ALGORITHMS

Gonzalez gave a simple 2-approximation algorithm for $k$-centre [2]. Ene et al. [1] developed a framework for adapting $k$-centre algorithms for MapReduce that, with high probability, gives a 10-approximation when implemented using a 2-approximation subprocedure. Malkomes et al. [3] recently gave an alternative 2-approximation scheme for $k$-centre that relates to a restricted case of our procedure.

Let there be $m$ machines, each with capacity $c$, and input set $V$ containing $n$ points. If $n \leq mc$ and $k\cdot m \leq c$, then Algorithm 1 runs in 2 iterations. The 4-approximation arises from implementing our algorithm with the 2-approximation algorithm of Gonzalez [2] as a subprocedure — the approximation ratio is $2i$, where $i$ is the number of iterations. When $km > c$, solutions can be obtained by running additional iterations, each adding 2 to the approximation factor.

Experimental results demonstrate that both parallel algorithms generally perform as well as the 2-approximation algorithm of Gonzalez, with our procedure performing up to 100 times faster than the alternatives. Our algorithm is designed to only use two passes under most conditions; the sampling algorithm uses at least four. For larger $k$, the Ene algorithm is theoretically equivalent to the Gonzalez algorithm, as shown in Figure 1(b) for $k = 100$.

Algorithm 1 MRG($V, k, m$)
1: while $|V| > c$ do
2: Arbitrarily partition $V$ into sets $V_1, \ldots, V_m$ such that $\bigcup V_i = V$ and $|V_i| \leq \lceil n/m \rceil$.
3: Send each set $V_i$ to a machine $p_i$.
4: In parallel, each $p_i$ runs a $k$-centre algorithm on $V_i$, and returns $S_i$, the $k$ centres found. Set $V = \bigcup S_i$.
5: Send $V$ to a single machine.
6: Run a $k$-centre algorithm on $V$; return the centres $S^*$.

$km > c$, solutions can be obtained by running additional iterations, each adding 2 to the approximation factor.

3. CONCLUSIONS

Our algorithm is simpler, faster, requires fewer iterations, and is able to find larger numbers of clusters than the alternative parallel algorithm of Ene et al. [1]. An extended version of these results, co-authored with Tony Wirth, will appear in ICPP’16.

4. REFERENCES

Privacy Protection of Large-scale Trajectories

[Extended Abstract]

Shuo Wang
Department of Computing and Information Systems
University of Melbourne
shuow4@student.unimelb.edu.au

Richard Sinnott
Department of Computing and Information Systems
University of Melbourne
rsinnott@unimelb.edu.au

ABSTRACT
The data analytics of the movement of people and their trajectories would significantly assist the improvement of policies for smart city management. However, with the global adoption of smart mobile devices equipped with localization capabilities and broad popularity of microblogging facilities like Twitter, the need for personal privacy has never been greater. This is especially so with computational and data processing infrastructures such as Clouds that support big data analysis. This study focuses on preserving social media (Twitter) users’ trajectories privacy. It proposes an innovative private trajectories calibration and publication model with differential privacy guarantees with enhanced utility.

CCS Concepts
• Security and privacy → Privacy protections
• Human-centered computing → Social media

Keywords
Calibration, publication, trajectories privacy, Social Media.

1. INTRODUCTION
Twitter has become a global phenomenon with over 400 million tweets made daily. Many users are unaware that often the geo-location of the tweet is also recorded, i.e. where they actually tweeted from and at what time they tweeted. With the extensive adoption of mobile devices with location awareness ability and the widespread popularity of location-based social networks (e.g. Twitter), users’ trajectories are being formed and captured at an exceptional rate. Generally, there are two essential issues that need to be realized: trajectory data contain private information of users that need data publication and analysis solutions with privacy guarantees, and the inherent heterogeneity of trajectories makes it inflexible to effectively perform trajectory analysis. Recently, differential privacy [1] has been widely used for the protection of location and trajectories data. The advantage of differential privacy for location privacy is that it allows to protect individual location information whilst still allowing the data to be used for analysis. Solutions that can limit the dangers of leaking location privacy would encourage more users to share their location information. But, the utility of solutions with differential privacy guarantees is constantly limited [2]. Therefore, it is essential to develop a trajectory calibration and publication approaches with differential privacy guarantees whilst enhancing the utility of such release data more generally.

2. DATA AND METHODS
The data collection is through the Twitter Search API and Streaming API, followed by data preparation, and the associated methods required to trajectories extraction and associated calibration and publication. These parts are implemented on the Australian National eResearch Collaboration Tools and Resources (NeCTAR) Research Cloud. Method used to generate a differential privacy-driven trajectories calibration and publication from raw geo-located Twitter trajectories data is achieved in four stages. (1) Boundary Insensitive Micro-aggregation (BIM) is proposed to create serial k-anonymity version of the location datasets that reduce the sensitivity of original datasets and thus decreases both the execution time and information loss. (2) An innovative feature-based clustering method with differential privacy guarantees (FBCM) is developed to find anchor points taking both density and important features of trajectory (e.g. direction change) into account. A private reference system is then built to calibrate original trajectories based on perturbed anchor points. (3) Trajectory calibration and sanitization system with different privacy guarantees (DPTCS) is proposed based on two-phases sanitizing solution to sanitize trajectories in largescale social networks. (4) Extensive experimental analysis on large-scale trajectories to assess the utility and effectiveness of our mechanism based on a set of evaluation metrics both on privacy and utility.

3. EVALUATION METRICS
The evaluation metrics used to measure the applicability of the approach described focuses on utility and privacy features of the Trajectory calibration and sanitization system with different privacy guarantees (DPTCS). These evaluation metrics contain several aspects: two effective utility evaluation metrics for anchor point, i.e. Summation of Distance Error and Recall to measure the information loss and the original data sets. And two effective trajectory similarity measures are proposed based on the geographic and semantic similarities.

4. REFERENCES
Constraint Programming Approach to Find Bounded-Distance Paths in a Graph

[Extended Abstract]

Diego de Uña, Graeme Gange, Peter Schachte, Peter J. Stuckey
Department of Computing and Information Systems
University of Melbourne
{diegod1@student., gkgange@, schachte@, pstuckey@unimelb.edu.au

ABSTRACT
Constraint Programming (CP) is a powerful technology used to model and solve combinatorial problems. Propagation algorithms are a key part of CP. In this paper we present our work involving a bounded path propagator that can be used to reduce the solving time of NP-hard problems where finding a path in a graph is part of the problem. Our propagator makes use of Lazy Clause Generation (LCG) to accelerate the search.

CCS Concepts
• Theory of Computation → Discrete Optimization;

Keywords
Constraint Programming; Path Algorithms

1. INTRODUCTION
Path-finding is an important task in (directed) networks. It arises in tasks such as graph layout, metabolic networks or collaborative path-finding in video-games among other examples. In many of these cases, though, side constraints make this problems highly combinatorial and no efficient algorithms exist. In this paper, we tackle these problems using CP, which provides a generic framework to model all sorts of NP-hard problems and solve them through a combination of exhaustive search and powerful propagation of inferences.

1.1 Constraint Programming
Constraint Satisfaction Problems (CSP) consist of constraints over a set of variables each with a domain set of possible values. A valid solution to a CSP is a valuation of these variables that satisfies all the constraints. Constraint programming solvers are used to solve CSPs. They interleave search and propagation. The former is simply an exhaustive search that tries to assign a value to each variable in turn until a conflict is detected or a solution is reached. The latter is a process by which propagators remove values from the domains of variables when they cannot be part of a solution given the decisions made so far. Propagators are thus simple algorithms that make inferences.

LCG [3] is a technique by which propagators generate Boolean clauses that capture the inferences they make. The solver can reuse these explanations to avoid repeating mistakes and accelerate the search. We call this learning.

1.2 Paths in Constraint Programming
Different ways of modelling paths in a problem exist. One could simply do a decomposition model of a path by using successor variables for each node in the graph or we could use specialised propagators that can do much more advanced reasoning. Previous approaches involved modelling paths as circuits [2] or with directed trees [1] (called PATH).

Our work improves on [1] by including learning in their tree propagator. On top of that, we build a bounded_path propagator that does the reasoning over distances.

The bounded_path propagator is composed of two levels. First, a fast lower bound (BPath) on the distance between source and sink is computed. If the distance between them is too high, we can prune the current branch of the search, thus reducing the search space. Second, we developed a Dynamic Programming algorithm (DPBPath) based on Dijkstra’s to compute shortest paths with mandatory nodes. This gives us a much tighter lower bound and prune the search space drastically.

In our propagator we also include learning; every propagation made is explained by a Boolean clause. We first developed explanations that could be computed fast based on edges that participate to short-enough paths (Expl). We introduced new minimal explanations that capture the entire power of the propagation in a clause (Expl*).

1.3 Results
Imagine a traveller has a long layover in a city and has a limited time to change airports. She can choose to visit some attractions in the city while on transit. Each point of interest (POI) has a reward and minimal time spent.

<table>
<thead>
<tr>
<th>Version</th>
<th>New York City (14 POI)</th>
<th>London (12 POI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Decisions</td>
<td>Time(s)</td>
</tr>
<tr>
<td>PATH</td>
<td>&gt;5034046</td>
<td>&gt;3600.00</td>
</tr>
<tr>
<td>PATH+BPath (Expl)</td>
<td>391746</td>
<td>379.19</td>
</tr>
<tr>
<td>PATH+DPBPath (Expl)</td>
<td>44606</td>
<td>48.82</td>
</tr>
<tr>
<td>PATH+BPath (Expl*)</td>
<td>361971</td>
<td>350.26</td>
</tr>
<tr>
<td>PATH+DPBPath (Expl*)</td>
<td>2690</td>
<td>37.45</td>
</tr>
</tbody>
</table>

Table 1: Profitable tourist path problem.
We see how out DPBPath algorithm alongside with the best explanations drastically improve the time needed to solve these problems.

2. REFERENCES
Effective Straggler Mitigation for Scalable Distributed Microscopic Traffic Simulation

[Extended Abstract]

Eman Bin Khunayn, Shanika Karunasekera, Kotagiri Ramamohanarao, Hairuo Xie
Department of Computing and Information Systems, The University of Melbourne
emanb@student.unimelb.edu.au, {karus, kotagiri, xieh}@unimelb.edu.au

ABSTRACT

Large-scale distributed simulations, which require periodic synchronization, often suffer from the straggler problem. In this work, we propose a novel parallel computational model called Priority Synchronous Parallel (PSP), and load balancing method to improve simulation performance through mitigating the straggler problem. Our approach is shown to achieve 80% performance speed up.

CCS Concepts

• Computing methodologies → Distributed algorithms;

Keywords

Distributed simulation; Synchronisation; Load balancing

1. INTRODUCTION

Large-scale microscopic traffic simulation (MTS) is useful tool to study and predict the detailed behaviour of traffic in big cities. Running such simulations require high computational power, which can be obtained through parallel and distributed computing. This is achieved by partitioning the simulation space into subregions to be executed in parallel on different processors. This parallel distributed implementation follows the Bulk Synchronous Parallel computational model (BSP) [3]. Each processor computes the calculations of assigned subregion locally and communicates with others to synchronise any needed information (e.g. moving vehicles) from other partitions, at the end of each simulation-step. However, implementations using BSP usually suffer from the stragglers problem, where the delay of any processor will slowdown the entire simulation. Straggler processors usually occur from imbalanced workload or communication delays among processors. In this work, we provide a new parallel computational model and load balancing techniques to reduce dependencies among parallel distributed processors to mitigate the straggler problem.

2. METHODS AND RESULTS

To improve the performance of MTS through addressing the straggler problem, we design a new parallel computational model called Priority Synchronous Parallel (PSP). This model is similar to BSP, however, it divides each simulation step into two phases alpha and beta. Alpha phase is where the high priority objects are computed and synchronised with dependant partitions (i.e neighbours), then the beta phase is for the rest of computations. The priority is measured based on the proximity of moving vehicles to the boundaries of processors partitions in a particular simulation-step. Furthermore, we developed a load balancing algorithm that is capable of balancing the workload and reduce the volume of data communicated among processors. The algorithm uses spatial features of simulated area and the layout of the road network in this area. In this algorithm, we transform the road network graph into a smaller graph called the processor graph. First, we use a grid to partition the road network space into cells, then we create a vertex in the processor graph for each cell in this virtual grid. Next, we add an edge between two vertices in the processor graph if there are any roads crossing between grid cells. Finally, we assign a weight to all vertices equal to the total length of all roads inside each vertex, and we assign a weight on edges based on the number of roads that cross between two vertices. Then, we use Multilevel graph partitioning algorithm (METIS) [1] on the weighted processor graph to achieve optimal workload balance that will maximise the area per processor and minimise the communication among processors. Our approach is implemented in SMARTS traffic simulator [2] and shown to achieve 80% performance speed up compared to using traditional BSP model (Figure 1).

Figure 1: Comparison of BSP vs PSP performance of running the simulation for Melbourne city with constant workload of 10,000 vehicles per processor

3. REFERENCES

Mitigating Impact of Short-term Overload on Multi-Cloud Web Applications through Geographical Load Balancing

[Extended Abstract]

Chenhao Qu  
CLOUDS Lab  
cqu@student.unimelb.edu.au

Rodrigo N. Calheiros  
CLOUDS Lab  
rcn@unimelb.edu.au

Rajkumar Buyya  
CLOUDS Lab  
rbuyya@unimelb.edu.au

ABSTRACT

One of the appealing features of cloud computing is its elasticity, which allows application providers to expand or shrink the amount of resources using auto-scaling. However, even if managed by an auto-scaler, sudden flash crowds and resource failures may still overload the application. With more cloud providers building geographically dispersed data centers all over the world, applications are commonly deployed in multiple data centers to better serve customers worldwide. In this type of deployment, instead of sufficiently over-provisioning each data center to prepare for occasional flash crowds or VM failures, it is more cost-efficient to over-provision each data center a small amount of capacity and balance the extra load among them when resources in any data center are suddenly saturated. In this paper, we propose a system that supplements and enhances state-of-the-art auto-scalers for applications deployed across multiple clouds. It is able to timely detect short-term overload situations and then autonomously handle them using geographical load balancing and admission control. We developed a prototype and evaluated it on Amazon Web Services. Evaluation results show that our system can timely detect overload events and effectively improve the application performance during resource contention periods.

CCS Concepts

- **Software and its engineering** → Cloud computing;
- **Networks** → Cloud computing

Keywords

Web Application, Geographical Load Balancing, Resource Overload, Auto-Scaling

1. INTRODUCTION

Cloud computing continues to gain rapid adoption for hosting web applications. One of its appealing features is its elasticity, which allows application providers to dynamically expand or shrink the amount of resources provisioned to their applications using auto-scaling. However, detecting workload changes and provisioning enough resources in cloud still demand considerable time (usually several minutes), which results to performance degradation and even unavailability of service during this period.

In web applications, it is common to observe a surge in request for services once in a while. This situation is called flash crowd and it can occur any moment with little or no warning. Cloud auto-scaler in these cases, cannot timely provision enough resources to deal with these situations. In addition, sudden failures of either software or virtual machines can also lead to application overload. Though failures are expected to be rare when the application is running on reliable resources and mature software, they may cause significant impact if the total resource pool is small. To make things worse, application providers now are deploying web applications on low cost cloud resources, such as spot instances [1], which makes the applications more prone to resource failures as such resources may be terminated by the provider when the market price exceeds the bidding price. Therefore, solely relying on auto-scaling is not enough to ensure high performance all the time and certain level of over-provisioning is required in production environments in preparation of these events even when the application is managed by an auto-scaler in the cloud.

Cloud providers have established their data centers all over the world, which enables their customers to deploy their applications in multiple geographically dispersed regions to better serve the worldwide population. As multi-cloud deployment is becoming more and more popular, we argue that in this type of deployment, when failures happen or flash crowd occurs to a data center, it is better to utilize the spare capacity already provisioned in other data centers to process as many exceeding requests as possible through geographical load balancing, instead of processing all the requests locally and degrading the performance of all clients, or rejecting the exceeding requests directly. This approach is viable as failures are unlikely to happen simultaneously in multiple data centers and flash crowds also seldom take place in a global scale at the same time due to culture and time difference.

In this paper, we propose a system that supplements and enhances state-of-the-art auto-scalers for applications deployed across multiple data centers. It aims to quickly detect and adapt to short-term overload caused by resource failures and flash crowds in each data center through geographical load balancing and admission control before the auto-scaler finishes provisioning new resources. Different from previous geographical load balancing solutions, our system relies on individual peer-to-peer agents deployed in the overloaded data center to realize fast and accurate geographical load balancing. During the overload situations, the agent deployed in the overloaded data center temporarily forwards certain amount of excessive requests to other data centers based on the proposed overload handling algorithm. We implemented a prototype and evaluated it on Amazon Web Services, who offer IaaS infrastructure in multiple geographically dispersed regions. Results show that our approach can timely detect short-term overload events and effectively improve the application performance during resource contention periods.

2. REFERENCES

Cost-efficient Processing of Large-scale Graphs on Public Clouds

[Extended Abstract]

Safiollah Heidari
CLOUDS Lab, CIS Department, The
University of Melbourne
sheidari@student.unimelb.edu.au

Rajkumar Buyya
CLOUDS Lab, CIS Department, The
University of Melbourne
rbuyya@unimelb.edu.au

ABSTRACT
Large-scale graph processing has gained attention during the past few years. As the world is going to be more connected by appearance of new technologies and applications such as social networks, Web portals, mobile devices and applications, Internet of things and so on, a huge amount of data are created and stored every day in the form of graphs consisting of billions of vertices and edges. In this paper, we present iGiraph, a cost-efficient Pregel-like graph processing framework for processing large-scale graphs on public clouds.

CCS Concepts
• Computer-communication Networks → Distributed systems;
• Graph Theory→ Graph algorithms; Network problems

Keywords
Cloud computing; large-scale graphs; big data; cost

1. INTRODUCTION
As Internet continues to grow, the world is becoming a more connected environment and the number of data resources is increasing beyond what had been predicted before. Amongst various data modeling approaches to store huge data, graphs are widely adopted to model complex relationships among objects [1]. A graph consists of sets of vertices and edges which demonstrate the pairwise relationship between different objects. Many applications and technologies such as social networks, search engines, banking applications, smart phones and mobile devices, computer networks, the semantic web, etc, are modeling and using data in the form of graphs [2].

As these graphs are growing bigger and bigger, traditional processing approaches are not useful and efficient anymore. So, distributed systems, and cloud computing in particular, can propose seamless approaches for large-scale graph processing. In this paper, we propose a cost-efficient graph processing framework called iGiraph

2. iGIRAPH
iGiraph uses a cost-efficient dynamic re-partitioning approach that utilizes network traffic message pattern to reduce the number of virtual machines (workers) during the processing by migrating partitions and vertices to minimize the cost. The new re-partitioning method also mitigates network traffic results in faster execution. To achieve these, we have combined different concepts together and proposed a novel graph processing approach:

1) We modified Giraph, which is an open-source graph processing framework that utilizes bulk synchronous parallel programming model. 2) We have used the concepts of inner vertices and border vertices in graph partitioning. 3) We used the concept of high degree vertices and introduced high degree partitions using that concept. 4) We differentiate between convergent and non-convergent graph algorithms for designing our cost-efficient framework. 5) We evaluated our approach on a cloud platform. The picture below shows the architecture of iGiraph.

3. EVALUATION AND RESULTS
We implemented our framework using NECTAR cloud and evaluated that by three different applications: shortest path algorithm, connected component algorithm and PageRank algorithm. We also used three datasets including YouTube, Amazon and Pokec. As our results show, iGiraph could significantly outperform Giraph by reducing the cost of VM utilization. This is because using convergent algorithms, the partitions shrink during the processing. So, iGiraph re-partition the graph after each superstep and terminates the VMs that are not needed. It has been shown that this method can even be executed faster than the method is using by Giraph and perform better.

On the other hand, for non-convergent algorithms, iGiraph does not terminate the VMs. Instead, it uses high-degree vertices to monitor the number of messages transferring between partitions and migrates high degree border vertices to the partition which has most of their adjacent vertices on it. Again here iGiraph can reduce the communication cost significantly compared to Giraph with even better execution time.

4. REFERENCES
Energy-efficient Scheduling of Cloud Application Components with Brownout

[Extended Abstract]

Minxian Xu
CLOUDS Lab
xianecisp@gmail.com

Amir Vahid Dastjerdi
CLOUDS Lab
vahid.av@gmail.com

Rajkumar Buyya
CLOUDS Lab
rbuyya@unimelb.edu.au

ABSTRACT

It’s common for cloud data centers meeting unexpected load-s like request bursts, which may lead to an overloaded situation and performance degradation. The popular solutions to solve this problem are abandoning some unresponsive requests or start more hosts to serve, but it’s not energy efficient to keep the hosts to be overloaded or turn on more hosts. This problem could be well solved by adding a feature named brownout in cloud applications or services, which either part of the optional components or some of self-contained microservices are deactivated, to reduce utilization for serving more requests and saving energy consumption. In this paper, we introduced a system model enabling brownout control at application component level, which could also be applied to microservices architecture. A scheduling strategy combining with brownout is also proposed, aiming at reducing Cloud data center energy consumption. The strategy has been implemented and tested in CloudSim and the results show that our newly proposed algorithm can reduce more than 25% energy, and there is a tradeoff between energy consumption, discount amount and the number of shutdown hosts. In addition, we provide several component selection polices for users to choose from.

CCS Concepts

•Software Engineering → Cloud Computing;

Keywords

Cloud Data Centers; Energy Efficient; Application Component; Microservices; Brownout

1. INTRODUCTION

Cloud computing provides compelling features such as adopting pay-as-you-go pricing model, lowering operation cost, highly scalability, easy access and reducing maintenance expenses. This makes Cloud computing attractive to business owners as it eliminates the requirement for users to plan ahead for provisioning, and allows enterprises to start from the small and request resource on demand. However, energy consumption by the cloud data centers has currently become one of the major problems for the computing industry. The growth and development of complex data applications have promulgated the creation of huge data centers, which heightened the energy consumption. Reducing server energy consumption comes to be a main concern of researchers.

Given the scenario that the budget and resource are limited, unexpected events such as overloaded tasks, may trigger performance degradation. Consequently, the application can be under saturated condition, in which some users cannot be served in a timely manner or experience latencies, some other users may not receive service at all [1]. The saturated applications would also bring over-utilized situation to hosts and contribute to high energy consumption.

It’s common that part of application components maybe not necessary for users, or it’s not necessary to provide all the containers or microservices of the whole service for the users. We investigate that it could be more energy efficient to downgrade user experience by disabling part of application components or microservices in service to relieve the over-utilized condition, thus serving more clients and reduce energy consumption.

Therefore, we take advantage of a paradigm called brownout. We are inspired from the concept of brownout in electric grids. The original meaning of brownout is the voltage shutdown to cope with emergency cases, in which light bulbs emit fewer lights and consumes less power. With brownout, under overloaded situation or high energy consumed situations, the optional application components could be deactivated to reduce utilization to serve more clients.

In this paper, we consider component-level control in our system model. The model could also be applied to container or microservices architecture. We model the application components as mandatory ones or optional ones. By deactivateing part of the optional components, the application utilization could be reduced and finally total energy consumption could be reduced as well. By appropriately disabling part of unnecessary functions, system could achieve better stability and less energy consumption. While under the market scenario, service provider may provide discount for users as part of the services are not provided.

Our objective is reducing energy consumption of data centers as well dealing with resource limitation and the revenues of service providers. Our contributions are: 1) We present an approach for reducing energy consumption through brownout, the experiments show that this approach could reduce more than 25% energy. 2) We investigate the trade-off between energy saving and discount amount, which assists service provider to design their policies with brownout according to their business objectives and preferences.

2. REFERENCES

Towards Performance-Oriented Deployment of Streaming Applications on Cloud

[Extended Abstract]

Xunyun Liu
The Cloud Computing and Distributed Systems (CLOUDS) Laboratory, University of Melbourne
xunyunl@student.unimelb.edu.au

Rodrigo N. Calheiros
The Cloud Computing and Distributed Systems (CLOUDS) Laboratory, University of Melbourne
rnc@unimelb.edu.au

Rajkumar Buyya
The Cloud Computing and Distributed Systems (CLOUDS) Laboratory, University of Melbourne
rbuyya@unimelb.edu.au

ABSTRACT
Stream processing is the ideal paradigm to enable fast aggregation and analysis of huge data volumes in motion. To continuously perform analytics on the fly within the dynamic stream, streaming applications have to be developed on top of Data Stream Management Systems (DSMS) and further deployed in a distributed computing environment for processing parallelization. In most cases, streaming applications along with the underlying DSMS are deployed on a cluster platform in which the deployment process of streaming applications is tuned towards the pre-configured environment. However, such platform-oriented deployment approach is outdated as we enter the cloud era and there are seemingly infinite resources to be provisioned and managed as needed. It is essential for the developers to have an empirical method to estimate resource usages and customize resource types for the special needs of continuous queries. In this paper, we propose a two-staged deployment framework that aims to achieve a pre-defined performance target with minimal resource costs. The first static planning stage would determine the initial deployment plan based on the performance profiling result with a proposed knapsack algorithm, while the later stage resolves the performance issues detected during the runtime through constant monitoring and iterative adjustments.

CCS Concepts
• Information systems → Data streams; Stream management; Database performance evaluation; • Social and professional topics → Quality assurance; • Software and its engineering → Software performance;

Keywords
Stream Processing, Data Stream Management Systems, Performance Optimization, Resource Management

1. INTRODUCTION
Stream processing is an emerging paradigm that complements the incompetence of batch model when the latency of processing is of vital importance. To achieve scalable and fault-tolerant processing over distributed resources, streaming applications are normally built on top of Data Stream Management Systems (DSMS), which provide the stream abstraction to conceal the underlying complexity of managing queues and organizing consumers for the upper level application logics.

State-of-the-art DSMSs, like Apache Storm, Samza, and Spark Streaming, have evolved to be highly scalable and distributed middleware in order to deal with the ever-increasing and ever-shifting processing needs. Within these systems, continuous queries of streaming applications are implemented as a series of operators standing on the continuous data flow. These operators, together with the oriented streams that interconnect them, compose a direct acyclic graph as the logical view of streaming application which is also called the application topology.

To enable fine-grained stream partition and parallelization, each operator normally consists of several replicas (aka tasks) that are designed to be distributed across the platform and get executed simultaneously. During the deployment process, it is the developers’ responsibility to guarantee that each operator would have sufficient number of tasks to handle its incoming workload, and each task would be properly allocated and scheduled in a way that resource contention caused by task collocation is mitigated and the overhead of inter-machine communication is reduced.

Though the amount of resources required is actually determined by the specific deployment process, the relationship between resource provisioning and other deployment decisions is still unclear [1]. Obtaining such knowledge would offer the opportunity to greatly reduce the cost of stream processing, such as decreasing energy consumption of a cluster by task consolidation, or saving monetary cost of using public cloud by provisioning as less resources as the performance indeed requires.

However, no existing work has investigated on performance-oriented deployment. There is no analytic model on resource provisioning that estimates and justifies resource needs for a certain streaming application to reach its performance target; there is also no auto-parallelization method for operators to decide their parallelism hints so that the application throughput would not be throttled by some operators lack of parallel tasks; Besides, though there are several advanced scheduling mechanism in the exiting literature to solve the task allocation problem, none of them is applicable to the scenario with such diversity that the amount of tasks and amount of underlying resources are both adjustable, which makes load balance and minimal communication overhead even hard to achieve among distributed nodes.

2. REFERENCES
Improving Spectral Based Fault Localization Using Simple Static Analysis

[Extended Abstract]

Neelofar
The University of Melbourne
neelofar.eme@gmail.com

Lee Naish
The University of Melbourne
lee@unimelb.edu.au

Rao Kotagiri
The University of Melbourne
kotagiri@unimelb.edu.au

ABSTRACT
Debugging is crucial for producing reliable software. One of the effective bug localization techniques is Spectral-Based Fault Localization (SBFL). It helps to locate a buggy statement by applying an evaluation metric to program spectra and ranking program components on the basis of the score it computes. We introduce a weighting technique by combining static and dynamic analysis and evaluate its performance both for small and larger programs. Results show that our technique improves the performance of a wide variety of fault localization metrics on single and multiple bug data.

CCS Concepts
• Software and its engineering $\rightarrow$ Software testing and debugging;

Keywords
spectral debugging, static analysis, dynamic analysis, fault localization metrics

1. INTRODUCTION
SBFL techniques using statements spectra are quite simple, using only four parameters extracted during program execution i.e. executed passed, not executed passed, executed failed, not executed failed [1]. However, their fault localization capability is hampered due to their limited information about the program. This research aims at improving the performance of SBFL techniques (dynamic analysis) by incorporating information gained by static analysis of program such as categorizing statements into different classes and estimating their likelihood of being buggy relative to other types of statements [3][2].

In this paper we make the following contributions:
• We investigate which category of program statements are more likely buggy. We focus on software programs written in C programming language and expect that these results equally applicable to other imperative programming languages.
• We use very simple static analysis to divide statements into different categories.
• We provide a simple generic modification to all metrics to exploit buggy behaviour based on the statement category.

<table>
<thead>
<tr>
<th>Without</th>
<th>With</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kulczynski2</td>
<td>21.2164</td>
<td>19.3385</td>
</tr>
<tr>
<td>Ochiai</td>
<td>21.5369</td>
<td>17.4380</td>
</tr>
<tr>
<td>Kulczynski1</td>
<td>22.7499</td>
<td>17.8934</td>
</tr>
<tr>
<td>Tarantula</td>
<td>23.0847</td>
<td>17.5743</td>
</tr>
<tr>
<td>O²</td>
<td>25.4841</td>
<td>21.9464</td>
</tr>
<tr>
<td>Zoltar</td>
<td>26.2260</td>
<td>23.2911</td>
</tr>
<tr>
<td>Ample</td>
<td>27.7600</td>
<td>24.2695</td>
</tr>
<tr>
<td>Wong2</td>
<td>34.5842</td>
<td>26.2914</td>
</tr>
<tr>
<td>Wong1</td>
<td>34.8494</td>
<td>25.9826</td>
</tr>
<tr>
<td>O²+</td>
<td>38.0182</td>
<td>31.8852</td>
</tr>
</tbody>
</table>

Table 1: Comparison of the average rank percentages of various metrics on multi-bug STS data with and without giving weights to the statements.

2. EXPERIMENTAL RESULTS
Table 1 shows results by applying our technique on multi-bug STS data. First column shows rank percentages without applying weights, while second column shows improvement in performance of different metrics in terms of lower rank percentages by applying our technique. A noticeable decrease in the rank percentages of the fault localization metrics shows that by categorizing and weighting statements before spectral based fault localization can improve the performance of fault localization metrics.

3. FUTURE WORK
The improvements in performance are particularly encouraging because the static information we have incorporated so far is very simple. With more static information incorporated, we would expect even greater improvements in performance. It is a common observation that deeply nested, complex and large statements have a higher chance of being buggy. We plan to incorporate this information in learning weights to further improve our technique as a future work.

4. REFERENCES
Combining BDD based Circuit Synthesis Technique with Masked Dual-Rail Pre-charge Logic to Eliminate Glitches in Circuits

[Extended Abstract]

Partha De
University of Melbourne
parthde@gmail.com

Udaya Parampalli
University of Melbourne
udaya@unimelb.edu.au

Chittaranjan Mandal
Indian Institute of Technology Kharagpur
chitta@iitkgp.ac.in

ABSTRACT

Repelling Power analysis attacks a workaround method is to randomize the intermediate results occurring during the execution of the cryptographic algorithm. Masking logic exploits this idea by ensuring that the power consumption of operations on randomized data is not correlated with the actual plain intermediate data. However, the various design models based on masking have neglected the possibility of multiple switching of the outputs of the gates in a single clock cycle although this phenomenon, termed as glitches, is typical for CMOS circuits. Binary Decision Diagram (BDD) based circuit synthesis mechanisms have been demonstrated to successfully safeguard secret information against PAAs by maintaining identical critical path lengths from the root to all leaf nodes, thereby ensuring that computation along each decision branch of the BDD will be through same number of transistors giving rise to identical delays. In this work, we demonstrate that this property of BDDs can be used to eliminate glitches in circuits.

CCS Concepts

• Security and privacy → Side-channel analysis and countermeasures;

Keywords

Power analysis attack; Binary Decision Diagram; Masking logic; Glitching; Voltage scaling; Leakage power.

1. INTRODUCTION

The basic idea behind this approach is that the power consumption of operations on randomized data should not be correlated with the actual plain intermediate data. Masking can be applied at two levels - algorithm level and gate level. Some algorithm level masking schemes can be found in [1, 2]. Masking at gate level was first proposed in [6]. Since masking at gate level leads to circuits where no wire carries a value which is correlated to an intermediate result of the implemented cryptographic algorithm, it is found to be more generic and therefore more amenable towards automation than masking at algorithm level. However, none of these masked based designing methodologies [1, 2, 6] considers multiple switching of the outputs of the gates in a single clock cycle which is typical for CMOS circuits. Therefore, glitching poses serious risk for circuits synthesized using any of the designing schemes. PAAs have been shown to correctly predict the secret data by analyzing the remnant glitches in CMOS circuits in [5]. To alleviate the problem of glitches, enhanced gate level masking approaches are proposed in [4, 7]. Moreover, the masked gates described in [6] are considerably large and hence more compact designs are possible using [4, 7]. However, in all these methods, the number of switching in a circuit is not always constant. Since switching power is a major shareholder of the total power consumption, the method of [3] employs a path balancing scheme for BDDs which ensures that all the paths from the root to a leaf node in a BDD passes through a constant number of nodes, thereby equalizing the number of transistor switchings and delay times for all possible input data.

In this work, we propose to combine BDD based circuit synthesis technique of [3] with masked dual-rail pre-charge logic of [7] to repel sophisticated PAAs that exploit EPE and/or glitches in a circuit by maintaining constant number of switchings along all possible paths between VDD to GND and applying a masking scheme for the gates. We have also added an extra stage in the synthesis process that employs voltage scaling and leakage power minimization to reduce the overall power consumption by the circuitry. Based on this ideas, we have designed a basic cell exhibiting the functionalities of AND, OR, NOT, NAND, XOR, etc.

2. REFERENCES

SLA-aware and Energy-efficient Dynamic Overbooking in SDN-Enabled Cloud Data Centers

[Extended Abstract]

Jungmin Son
Cloud Computing and Distributed Systems Lab
ejungmins@student.unimelb.edu.au

Rajkumar Buyya
Cloud Computing and Distributed Systems Lab
rbuyya@unimelb.edu.au

ABSTRACT
Power management of cloud data centers received great attention from industry and academia since their operational cost is expensive due to their high energy consumption. Resource overbooking is one method to reduce the usage of active hosts and networks by placing more requests to the same amount of resources. Network resource overbooking can be provided by Software Defined Networks (SDN) that can consolidate traffics and control Quality of Service (QoS) dynamically. However, previous approaches employed fixed overbooking ratio to decide the amount of resources to be allocated, which in reality may cause excessive Service Level Agreements (SLA) violation with unpredictable workloads. In this paper, we propose dynamic overbooking strategy which jointly leverages virtualization capabilities and SDN for VM and traffic consolidation. With the dynamically changing workload, the proposed strategy allocates more precise amount of resources to VMs and traffics. This method can increase overbooking in a host and network while still providing enough resources to minimize SLA violations. Our approach calculates resource allocation ratio based on the historical monitoring data from the online analysis of the host and network utilization without any pre-knowledge of workloads.

CCS Concepts
• Networks → Cloud computing; Data center networks;

Keywords
Cloud computing; data center network; Software-Defined Networking; resource provisioning

1. INTRODUCTION
Providing subscription-oriented cloud computing services has attracted great deal of attention in both industry and academia. One of the major concerns in cloud computing is tremendous electrical power consumption in cloud data centers. According to the US Natural Resources Defense Council [1], data centers in the U.S. consumed about 91 billion kilowatt-hours of electricity in 2013, which is roughly twice of the electricity consumption in New York City. Moreover, the electricity consumption of data centers is subjected to rise to about 140 billion kilowatt-hours annually until 2020.

Over-provisioning of resources (hosts, links and switches) is one of the major causes of power inefficiency in data centers. As they are provisioned for peak demand, the resources are under-utilized for the most time. For example, the average utilization of servers reported to be between 10-30% for large data centers [3], which results in a situation where considerable capacity of data center is idle. Similarly, provisioning of network capacity for peak demand leads to energy waste, which can be reduced through the effective use of Software-Defined Networks (SDN). With SDN, now cloud data centers are capable of managing their network stack through software and consider network as one of the key elements in their consolidation techniques.

While overbooking strategies can save energy, they also increase the chance of SLA violation when either host or network is overloaded. If the consolidated VMs or traffics reach the peak utilization at the same time, insufficient amount of resources would be allocated which will delay the workload processing. The main objective of our approach is to ensure both SLA satisfaction and energy saving without compromising one for the other. We aim to reduce SLA violation rate while increasing energy savings.

We propose dynamic overbooking algorithm for joint host and network resource optimization that, in comparison to previous works, has three novelties. Firstly, our approach employs a dynamic overbooking strategy that dynamically adapts to the workload instead of using a fixed percentile. Secondly, it is designed to work without the prior knowledge of the workload. Lastly, we consider initial placement and consolidation strategies together to find the most effective combination for energy saving and SLA satisfaction. The proposed algorithm is implemented in CloudSimSDN [2], a cloud simulation tool supporting various SDN features such as dynamic network configuration and programmable controller.

2. REFERENCES
Semantic-aware Service Definition and Discovery in IoT Using Linked Data and CoAP

[Extended Abstract]

Farzad Khodadadi
Department of Computing and Information Systems, University of Melbourne
fkhodadadi@student.unimelb.edu.au

ABSTRACT
The future of machine-to-machine (M2M) communications relies on reusability, scalability, and interoperability of services defined by each involved entity. Internet of Things (IoT) aims to provide an environment where smart devices can easily expose their services, while providing accurate service discovery functionality for clients to consume them. In this paper, we leverage the Constrained Application Protocol (CoAP) protocol and our defined ontologies to semantically represent services through modern service presentation formats such as JSON-LD and API definition languages.

CCS Concepts
Information systems → Information retrieval; Information systems → Service discovery and interfaces

Keywords
Internet of Things, linked data, service discovery, semantics, ontology

1. INTRODUCTION
Service-oriented architectures and methodologies have been widely adopted and studied in distributed systems, well before the emergence of IoT. However, due to huge numbers of entities and respectively, large service pools in IoT, the trend has shifted towards using more lightweight services [1]. Traditional SOAP-based services have been gradually replaced by RESTful services and APIs are now the new players in this field. Briefly, APIs are easier to define, invoke, share, and monitor compared to other service definition methods and enterprises can commercialize their APIs more effectively, since more tools and documentations are available for end-user consumers [2].

Generally, IoT environments span across multiple domains and to make interoperability between different domains functional, one suggested approach is using semantic technologies. To achieve this goal, we need to combine the power of semantic-aware service discovery with modern and standard service definition methods that take into account IoT’s special requirements such as limited and scattered resources. In this paper, we propose a system that facilitates service discovery by leveraging API definition languages and implements an efficient discovery component by semantically tagging services. Furthermore, it can use both CoAP and HTTP requests for service discovery and semantics storage.

2. SYSTEM ARCHITECTURE
Every object in a typical IoT environment has characteristics and services that need to be defined and stored before they can be used. According to our architecture, each “Thing” can be defined using a JSON-LD file that maintains a unique id for its particular domain, alongside with a list of properties. These properties can range from device’s specific characteristics like hardware capabilities to services it offers which are defined using an API definition language such as Swagger. RDF and JSON-LD have been developed for machine-readable linked data and classic Web services.

Service discovery is our main module here and to implement it, we first need to semantically annotate services and properties according to an ontology derived from Minimal Service Model (MSM) [3]. The aforementioned ontology has been extended to include SSN-XG, which is a widely adopted ontology for defining sensor properties. Semantic data is stored in RDF format in an Apache Triple Database and a SPARQL endpoint is provisioned to query it. Additionally, Things Description Documents represented in JSON-LD format are stored and indexed by Apache Solr for fast retrieval.

3. COAP INTEGRATION
Constrained Application Protocol (CoAP) is a lightweight UDP-based protocol developed with constrained devices in mind that helps reduce the overhead of using normal HTTP request and responses. One benefit of using CoAP is that IoT devices will consume less power and achieve better performance.

To support CoAP-enabled devices in our proposed service discovery system, a proxy server is introduced to handle communications that use this protocol. The smart proxy knows which device is CoAP-ready and chooses the right communication protocol according to each device’s capability.

4. PERFORMANCE EVALUATION
According to our findings, using CoAP instead of normal HTTP methods reduces the communication cost when lots of service calls are made by IoT entities. Furthermore, addition of semantic tags to services that can be invoked using CoAP has improved the accuracy of our proposed service discovery component.

5. REFERENCES
An Online Algorithm for Task Offloading in Heterogeneous Mobile Clouds

[Extended Abstract]

Bowen Zhou
Cloud Computing and Distributed Systems Laboratory, University of Melbourne
bowenz@student.unimelb.edu.au

Rajkumar Buyya
Cloud Computing and Distributed Systems Laboratory, University of Melbourne
rbuyya@unimelb.edu.au

ABSTRACT
In this paper, we address mobile code offloading and scheduling problem as an optimization problem in a heterogeneous mobile cloud environment. We propose a context-aware mixed integer programming formulation for scheduling mobile tasks on the local device or offloading it to other computing resources in the HMC. The objective is to minimize the global task completion time (i.e. makespan). To solve the problem in real-time, we propose a deterministic online algorithm based on the ski-rental framework with the competitive ratio of $2 + \frac{\alpha}{2}$, where $\alpha$ is the percent of task execution time considered in the overall cost by the scheduler in HMC.

1. INTRODUCTION
As mobile computing technologies developed enormously in recent years, mobile applications such as cognitive applications (e.g., optical character recognition, face detection) and augmented reality are gaining popularity among mobile device users. These applications typically require intensive computation and incur considerable energy consumption. However, compared to desktop computers, mobile devices still provide a relatively inferior performance in terms of processing capacity, memory, storage capacity and battery life to support long-time services. As a result, the gap between resource-constrained mobile devices and computation intensive applications has posed a significant challenge.

Mobile cloud computing (MCC) is introduced by bringing the benefits of cloud computing to augment the resource-constrained mobile devices. Code offloading provides an approach that migrates computation intensive tasks to cloud in order to enhance the processing capacity and reduce the energy consumption of the mobile device.

The heterogeneous mobile cloud (HMC) environment contains different types of computing resources such as public clouds, private clouds, cloudlets [2], and mobile ad-hoc networks (MANET) in proximity that can be utilized to offload mobile tasks. Each mobile device within the shared environment is able to offload its mobile tasks to other available machines, and vice versa. For the individual mobile device, it is important to make the offloading decision based on network conditions, load of other machines, and mobile device’s own constraints (e.g., mobility and battery). Moreover, to achieve a global optimal task completion time for tasks from all the mobile devices, it is necessary to devise a task scheduling solution that schedules offloaded tasks in real time.

2. RESEARCH METHOD
We jointly study the mobile code offloading problem and task scheduling problem for offloaded tasks as a mobile code offloading and scheduling problem (MCOSP). The heterogeneity in both tasks and machines makes MCOSP an NP-hard problem [1].

To solve the problem, first, we propose the models of computation, energy consumption, monetary cost and time-to-failure for mobile devices that represent the heterogeneity of the proposed heterogeneous mobile cloud environment. Second, we formulate the MCOSP as a mixed integer nonlinear programming (MINLP) problem based on the models of HMC and provide an analysis on its hardness. Then we transform it into a mixed integer linear programming (MILP) formulation using linearization and solve the problem using the branch-and-bound algorithm. Finally, to provide a practical solution for the problem, an online real-time scheduling algorithm for MCOSP based on ski-rental framework is proposed to obtain competitive near-optimal schedules for large workloads. Experimental results shows that the proposed online algorithm generates around 4-competitive makespan on average comparing to offline optimal solution in terms of scheduling performance, and scales well in terms of offloading gains when the number of mobile users increases in the HMC environment.

3. REFERENCES
People and Organisations
Exploring information artefacts of person centred discharge planning (extended abstract)

Nyree J. Taylor  
M (Phil) Engineering  
Nyreet@student.unimelb.edu.au

A/P Reeva Lederman  
Department of Engineering, CIS  
Reeva.leaderman@unimelb.edu.au

Dr. Rachelle Bosua  
Department of Engineering, CIS  
Rachelle.bosua@unimelb.edu.au

Abstract

Little is known about the artefacts that help inform technology solutions to support people returning home following hospital stay. Content, delivery, timing and unique personal circumstances form integral components to human computer interactions and user centred design. This paper explores how person centred artefacts can be used to inform information technologies to refine and more appropriately target person-centred discharge information, thus facilitating transitions from hospital to home.

CCS Concepts

• Human computer interaction → HCI design and evaluation methods → User models, user studies, heuristic evaluation.

Keywords

Information technology; discharge planning; information needs; patient centred information.

1. Introduction

There is increasing pressure on hospitals for better service delivery and shortened length of stays; much of patient recovery occurs at home[1, 2]. To facilitate continued self-care, useful person-centred discharge information is required for patients to self-manage and take care of their own recovery in the home.

2. Background

Fostering independence is a key component of effective and efficacious discharge planning. It’s important that patients are discharged with sufficient, timely and adequately-presented discharge information enabling them to manage their own care once home. Existing information, available online, in paper formats and through media is often voluminous, overwhelming and not clearly targeted [3]. This project is the first of a two-stage study to investigate how, when and to what extent patients use information provided to them upon discharge. Outcomes from the project can be used to help inform patient-hospital communications to better support patient safety and maintain sustainable health care routines once patients have been discharged.

3. Research Question

The study is focused on addressing the research question: “How can information technology be used to present person centred discharge information to best assist patients’ transition from hospital to home?”

4. Theoretical Foundation

A combined lens of Roy’s Adaptation Model (RAM) and Adaptive Structuration Theory (AST) forms the theoretical underpinnings of this study and seeks to explain the interplay between humans, their immediate environment, and the role of nursing information at discharge[4, 5]. Through the establishment of a conceptual framework (Figure 1), this theory combination seeks to understand person-centred artefacts including preferences, needs and values for discharge information and dependent outcome constructs from AST.

5. Design

An exploratory research design is used to examine information artefacts in discharge information for patients. Data collection instruments will be interviews and observations.

Participants will be female, aged between 30-59 years with admission to hospital for minor cardiac procedure. Cardiovascular disease is the biggest killer of Australian Women and this age group has been identified as a next wave of population growth [6, 7]. Females have been purposefully selected to seek to understand how hospitals acknowledge and incorporate a women’s personal circumstances into the pre/post discharge information.

6. References

Increasing Clinical Data Entry Quality Through Automated Web-based Feedback

[Extended Abstract]

Stephan Glöckner
Department of Computing and Information Systems, University of Melbourne
sglockner@student.unimelb.edu.au

Prof. Richard O. Sinnott
Department of Computing and Information Systems, University of Melbourne
rsinnott@unimelb.edu.au

ABSTRACT

Manual data entry is still used in the majority of clinical trials worldwide. However, entering data manually leads to low data quality. This is primarily caused by the motivation of data entry personnel. To improve this we introduce a web-based data quality feedback mechanism as mechanism to motivate users to enter higher quality data. Results of our case study confirms this thesis, but needed to be extended to gain more insights into intrinsic/extrinsic motivational factors.

CCS Concepts

- Social and professional topics ➔ Quality assurance
- Human-centered computing ➔ User studies

Keywords

Data Quality Feedback, Data Entry, User Motivation

1. INTRODUCTION

Clinical research captures trial information either electronically (EDC) or manually (MDE), often into a centralized trial registry. MDE can deliver more complete data, but it can be more inaccurate compared to EDC. Accuracy and completeness are only two dimensions of data quality. The gold standard for improving data quality is double data entry. But this method is highly inefficient with high costs and time vs. ~3-5% quality improvement. Other, more efficient methods for data quality improvements are needed.

Case studies have shown that feedback on data contribution leads to an increase in the motivation for data entry personnel [1]. We argue that more participation and encouragement to enter data into clinical registries also increases data quality.

For this, we introduced a model for web-based automated data quality feedback (DQF) that assists user during data entry. DQF to the investigator already showed data quality improvements of 13% within 6 months [2].

2. METHOD

To prove the efficiency of real-time DQF and community DQF we developed a stage-wise observational study to assess data completeness (DC) trends and the impact of DQF categories (see Table 1).

Table 1. Study Design - 2 Stages of DC Feedback

<table>
<thead>
<tr>
<th>Stage</th>
<th>Added records in Stage II</th>
<th>DC trend Stage II vs. Stage I</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>0</td>
<td>+13.33 %</td>
</tr>
<tr>
<td>II</td>
<td>15</td>
<td>-40.40 %</td>
</tr>
<tr>
<td>III</td>
<td>1</td>
<td>-11.73 %</td>
</tr>
<tr>
<td>IV</td>
<td>0</td>
<td>+0.70 %</td>
</tr>
</tbody>
</table>

3. RESULTS

DC was analyzed after 3 months of Stage II. Results showed increase of DC in four of six centres. IEDU1 showed a significant increase of DC in Stage II by improving already existing records. ITUD2 had a dramatic fall of DC, caused by entering new incomplete records into INPDR (see Table 2).

Table 2. DC Stage II vs. Stage I

4. CONCLUSIONS

Results showed an increase of DC of existing records, but a larger sample size is needed. To gain further insights into the DQF and users intrinsic (self-motivation – Stage II/III) and extrinsic motivational factors (community competition – Stage IV), the study is to be extended with two additional stages (see Table 3).

Table 3. extended Study Design

Stage III Stage II feedback with users average DC
Example Feedback: “Record has been updated with 30% DC, but your average DC is 55%.”

Stage IV Stage III with community average DC
Example Feedback: “Record has been updated with 30% DC, but your average DC is 55%, colleagues scored an average DC of 70%.”

5. ACKNOWLEDGMENTS

Our thanks to INPDR for funding this project and to Mr. William Hu for the software development.

6. REFERENCES


Understanding the Role of Bodily Communication in Video Consultations for Physiotherapy

[Extended Abstract]

Deepti Aggarwal
Microsoft Research Center for SocialNUI
University of Melbourne, Australia
daggarwal@student.unimelb.edu.au

Abstract
This research explores the role of bodily communication in video based clinical consultations for physiotherapy. Through an ethnography-informed study, I discuss the challenges that clinicians and patients faced during video consultations.

CCS Concepts
• Computer Interaction; Field Studies • Empirical Studies in collaborative and social computing

Keywords
Video-mediated communication; Clinical Consultations; Bodily Communication

1. INTRODUCTION
Bodily communication [2] is defined as a crucial aspect of clinical consultations. Example of bodily communication includes facial expressions, gaze, hesitations, and touch. In a clinical setting, bodily communication is used to establish rapport, to articulate the health issue and to suggest the treatment during a clinical discourse [4]. For example, a physiotherapist pays attention to the patient’s distorted and careful movements to check patient’s recovery. Similarly, a psychiatrist makes use of the patient’s abnormal body language, pale face, and hesitation in speaking to understand their stress level.

This research examines the significance and challenges of bodily communication during video-based clinical consultations for physiotherapy. Previous literature on video communication in non-clinical settings suggest that certain bodily information such as hand gestures, eye gaze and spatial orientation gets missed when we move our conversation from physical to digital space [5]. However, such bodily cues are always emphasized as crucial part of clinician-patient interactions during face-to-face consultations. This raises the need to understand whether video technology can support the essential bodily information required to accomplish different clinical tasks. Such an understanding is required to ensure that the introduction of video technology does not hinder the specific needs of clinicians and patients, and to guide the design of future technologies that could further enhance the clinician-patient interactions.

2. STUDY DESIGN
I conducted an ethnography-informed field study at the Pain Management Department of Royal Children’s Hospital. Over a period of 8-months, I observed 10 consultations: 7 video and 3 face-to-face. Observing face-to-face consultations provided me with background knowledge of the key aspects of clinical consultations for physiotherapy, which I utilized to understand the strengths and limitations of video consultations. Participants of the study included 2 physiotherapists at the hospital who recruited 5 patients suffering with chronic pain disorders in different body parts. Apart from observations, I also utilized semi-structured interviews and informal discussions to collect rich data.

3. FINDINGS
The study highlights that the clinicians rely upon a wide range of bodily cues related to body movements, postures, and tactile aspects of the patient’s body (for more details, refer [1]). Clinicians observe the patient’s bodily cues right from the start of the consultation, to get a complete picture of their health. Additionally, different bodily cues are relevant across different phases of the consultation (defined by [3]): Opening, History Taking, Examination and Diagnosis, Treatment and Closing. These bodily cues are naturally available in face-to-face consultations; however, some of these cues get missed over video. The lack of these bodily cues limited the information space of the clinicians. Consequently, clinicians relied more on verbal elaboration of the patient, which did not fulfill the needs aptly.

The insights gained from the study have opened up four design dimensions that speak to: 1) accommodating asymmetries of the roles and responsibilities underlying in the clinical setting; 2) expanding the perspectives of video consultations both in terms of field view and time; 3) augmenting visual acuity with computational (non-visual) technologies; and 4) using sensing technologies to communicate essential tactile information.

4. CONCLUSION
This study indicates that despite the challenges of communicating bodily cues during video consultations, clinicians find video consultations beneficial, particularly for follow-up consultations as it save patients trips to the hospital and their disruptions to schooling. Future video consultation systems should be designed to support essential bodily communication.

5. REFERENCES
Observing boardgame play in a research setting
[Extended Abstract]

Melissa J. Rogerson
Microsoft Research Centre for Social NUI
melissa.rogerston@unimelb.edu.au

Martin Gibbs
Microsoft Research Centre for Social NUI
martin.gibbs@unimelb.edu.au

Wally Smith
Microsoft Research Centre for Social NUI
wsmith@unimelb.edu.au

ABSTRACT
In this paper, we describe the use of a laboratory study as a novel and relevant research method for studying the play of tabletop boardgames. We present preliminary results from a study of boardgame play using a wearable eye tracker, which demonstrates its utility and application in this research context. Further, we examine the applicability of whole-of-room recording and gaze tracking for this type of research in non-digital environments.

CCS Concepts
Human-centered computing → User studies

Keywords
Boardgames, play, interaction, gaze, eye tracking, games.

1. INTRODUCTION
For hobbyist boardgamers, playing boardgames is an inherently material practice [5] which is not restricted to the home but may occur in different spaces and locations [4]. To the hobbyist, the critical factors – beyond the availability of other players – are afforded by the space itself: its light and ventilation, minimal noise, availability of an appropriate table and chairs [5].

To study play of boardgames, the researcher can either go to the participants – observing play in a private home or in a public setting such as a café, library [2] or game club – or provide an appropriate space where play can occur. In this paper, we describe the use of a user experience laboratory as a gaming environment to enable play and observation to occur concurrently. The laboratory offers the researcher the opportunity to record the play session and observe it remotely, without joining the participants in the room. Further, it enables innovative use of HCI technologies to more directly understand the play experience.

We propose a novel method combining whole-of-room video recording and eye gaze tracking glasses to study the play of a traditional form boardgame around a table. We present some preliminary recommendations for the types of data that can be obtained from the glasses, as well as the types of game situation that are best suited to longitudinal and consolidated analysis.

2. METHOD
Participants for this study were recruited using a snowball method, initiated through posts on the first author’s social media accounts and messages to local gaming groups. Volunteers completed an online survey indicating their age, gender, gaming experience and availability. Following this, they were invited to attend a two-hour session in the Interaction Design Laboratory at The University of Melbourne between March and June, 2016. To date, we have conducted six sessions with 30 participants, who have played a total of eleven games (seven unique).

Sessions were structured to allow for a brief initial welcome as well as a final (15 minute) focus group discussion, with a small game library available for play. One participant in each session wore a Tobii Pro Glasses 2 eye tracker, which monitored and recorded their gaze throughout the session. At the end of each session, a brief focus group collected participant feedback.

3. FINDINGS & CONCLUSION
Although eye tracking has been widely used in interaction design research on computer monitors, the use of glasses that can track a user’s gaze in non-screen-based environments is a more recent development. Gaze tracking has been used to study a range of non-screen situations including detecting eye contact, monitoring driver vigilance and in virtual space. [1; 3; 6]. Gaze data enables rich and deep analysis both of individual sessions and of amalgamated data, but there are limitations both in how it can be used and in how it can be consolidated. Whilst it is relatively simple to analyse screen-based gaze data using heatmaps and occlusion maps, the ‘open world’ approach enabled by gaze glasses uses a much broader space and is correspondingly more complex to analyse.

Our analysis and observations to date suggest that video recordings from the eye tracker are particularly compelling and interesting in this research context. These recordings, supplemented by the whole of room video recordings and focus group feedback from players, show the interesting decisions that players face during their turn and demonstrate their sustained attention during the game. Heatmaps, however, appear to be of more limited use, at least in this relatively open setting. Although they amalgamate session data and highlight key activity areas, further work is needed to establish their application.

4. ACKNOWLEDGMENTS
We gratefully acknowledge the support of the Microsoft Research Centre for Social NUI.

5. REFERENCES
Technology at Mealtimes: The Good, the Bad, and the Ugly

Hasan Shahid Ferdous
Microsoft Research Centre for Social NUI, Dept. of CIS, University of Melbourne
hferdous@student.unimelb.edu.au

ABSTRACT
We investigate the familial interactions with technologies around domestic mealtimes. We seek to understand how technologies are used and negotiated amongst family members and the influence of technologies on the content and context of their interactions. We focus on the potential scope for interventions in the social and material configuration enacted through the presence of smart-devices at the dining table. We present a novel system ‘TableTalk’ and discuss how a detailed understanding of commensality informs the design of new technologies and how such technologies might reconfigure our practices of shared mealtimes.

CCS Concepts
• Human-centered computing → Empirical studies in collaborative and social computing;

Keywords
Family; collaborative use; mobile devices; commensality.

1. INTRODUCTION
We explore the relationship between technology and everyday commensality. Alongside the act of eating together, it alludes to shared dependency, accountability, storytelling, planning, socializing children, and other social, political, and cultural aspects of the mealtime experience. Hupfeld et al. [1] provided a detailed account of the everyday practices associated with family meals and how it relates to the ecology of mealtime artifacts and spaces. We extend this by exploring the ways in which a broader set of everyday technologies become implicated in the social configuration of commensal practices and family relations at mealtime. Rather than making strong moral arguments with respect to the position of technology within mealtime behavior, we look to the ways that such technologies contribute or detract from any idealized notions of family order in these settings.

2. FINDINGS
Our exploration of the current practices of technology usage during family meals shows how technology is integrated in a dynamic and situated fashion into the mealtime activities. We reveal certain circumstances in which background technologies come to the foreground, visible devices are hidden, unwanted distractions become desired, and ordinary technologies support celebratory occasions. Technology often served as a conversational resource during mealtimes - it was, in itself, a topic of conversation.

Figure 1: TableTalk brings devices together to symbolize the communal aspects of commensality.

To this end, we present TableTalk, an application that integrates the personal smart-devices of mealtime guests into a single shared display (Fig. 1). The aim here is to enable the sharing of personal content (e.g., photos, music, social media posts) for the collaborative construction of a mealtime experience. TableTalk both embodies elements of commensality in its design (e.g., conversation, accountability, etc.), and also supports commensal experiences through its design (e.g., togetherness, shared reminiscence, etc.). Through a field deployment study, we derive practical insights into how digital technology can be designed as a part of commensal experience to support and configure shared space, data, narratives, and subsequently, interactions. Specifically, we consider the implications of the design on commensality in the (1) physical setup of devices and choice of personal content, (2) familial interaction with and around the technology, and (3) impact upon various aspect of everyday commensality. Finally, we offer novel understandings of the potential role of technology for commensality in the family home.

3. CONCLUSION
We explore current practices around the technologies and analyze how they support familial conversation, provide relaxation, and enable celebration, albeit not without occasional tension. Overall, our study demonstrates that through sensitive design, deployment, and recognition of family norms, expectations, and responsibilities, technology designers can support and enhance interaction at family mealtimes.

4. REFERENCES
Information Security Management Practices in Organisations

[Extended Abstract]

Moneer Alshaikh
Department of Computing and Information Systems
The University of Melbourne
Malshaikh@student.unimelb.edu.au

Sean B. Maynard
Department of Computing and Information Systems
The University of Melbourne
Sean.Maynard@unimelb.edu.au

Atif Ahmad
Department of Computing and Information Systems
The University of Melbourne
Atif@unimelb.edu.au

Shanton Chang
Department of Computing and Information Systems
The University of Melbourne
shanton.chang@unimelb.edu.au

ABSTRACT
There is growing recognition of the role that management performs in protecting organisational information. However, our review of the academic and professional literatures did not find an empirically sound and coherent view of the range of management activities that can be applied as part of an information security program. As a result, organisations have insufficient guidance on what methods can be implemented to meet security objectives. Further, organisations have no empirically evidenced benchmark against which management practices can be assessed. This research project aims to develop a rigorous, comprehensive and empirically evidenced model of information security management practices (ISMPs) to provide organisations with comprehensive guidance.

CCS Concepts
• Information systems → Security and Protection engines; Management of computing and Information Systems

Keywords
Information Security management, Policy, Incident Response, Risk Management, Security awareness

1. INTRODUCTION
The threat of leakage of trade secrets and intellectual property, disruption of mission-critical systems, and malicious attack from both insiders and outsiders makes information security a high priority for organisations. Although technical security controls have always played a critical role in reducing security risk exposure, recent research has highlighted the critical role of managerial controls in the pursuit of security objectives (e.g. see Knapp & Ferrante 2012; Ahmad et al. 2012; Lim et al. 2012).

Prior research in information security management (ISM) focuses on a variety of issues such as: the importance of security policy, development of good security policy, use of deterrence theory to investigate employees’ compliance with security policy, risk management, incident response and security awareness. However, it is argued that much of this research is disjointed and does not provide comprehensive guidance to organisations about information security management practice (Qingxiong et al. 2008, Lim et al. 2012).

The lack of a comprehensive model that provides sufficient guidance on the range of managerial practices of information security has serious consequences for the modern organisation given the significant exposure to security threats. Organisations embarking on an ISM program have insufficient guidance on what methods can be implemented to meet security objectives. Further, organisations have no widely accepted benchmark against which improvement in existing management practices can be assessed.

2. Aim
This research project aims to develop a model of ISMPs to address the need to identify, classify and understand the management practices of information security. The model will provide organisations with a means to understand, implement and assess ISMPs in a systematic and comprehensive manner in order to enable them to protect themselves from a wide range of threats.

3. Research Design
This research is exploratory and follows a qualitative research design. A preliminary model of ISMPs has been developed from the literature. The model is to be refined and validated by information security experts. Thirty semi-structured interviews with information security experts (e.g. security managers, chief information security, security consultant) have been conducted. Data collected from semi-structured interviews, is currently being analysed using open, axial and selective coding as described in Neuman (2006).

4. REFERENCES
The immobility of mobile devices: insights from video-mediated sessions with ethnic communities in Kenya

[Extended Abstract]

Kagonya Awori
Microsoft Research Social Natural User Interfaces Lab, University of Melbourne
sawori@student.unimelb.edu.au

ABSTRACT
Recent socio-cultural movements have motivated an interest from diaspora youth in developing their cultural identities at their places of migration. In this paper, we describe one of our findings from a field study with ethnic communities in Kenya, where we carried out 11 video-mediated sessions where elders, on the homelands, taught practical indigenous knowledge skills to youth in the diaspora. Learnings from this study will enable the design of a prototype that allows the unique social, physical and situated interactions in indigenous settings, to be performed when participants are remotely located.

CCS Concepts
• Human-centered computing ➔ Web-based interaction;
• Hardware ➔ Displays and imagers

Keywords
Remote collaboration; indigenous knowledge; mobile devices; video-mediated interaction

1. INTRODUCTION
Recent socio-cultural movements, such as decolonisation (e.g. among Kenyan intellectuals [1]); pan-Africanism (e.g. among Ghanaian diaspora [2]); and transnationalism (e.g. among the Chamorro in North America [3]), have motivated interest in diaspora populations, in developing their cultural or ethnic identity at their place of migration. Likewise, our previous research has noted an interest from diaspora youth in reviving, practising and fostering their ethnic or indigenous culture at their place of migration [4]. Given this interest from the youth, our work focuses on designing a video-mediated learning environment where elders living in the countries of origin, can teach practical cultural skills and knowledge to youth living in the diaspora. We carried out 11 live video mediated sessions between elders in rural Kenya and youth in the diaspora. The purpose of this study was to uncover what social, physical and cultural interactions need to be supported when elders engage with the remotely located youth when teaching practical indigenous skills over live video sessions. By uncovering the interactions that were not adequately supported in these learning sessions, we were able to identify what gaps can be filled further by technology. The next section talks about one of our main findings.

2. DISCUSSION
2.1 The challenge of mediating movement and mobility over distance
In our study, the elders used Skype and an iPad to interact with the remote learners. However, we observed that despite the mobility of an iPad, elders found it troublesome to manage the device and also teach practical indigenous skills and practices at the same time. Also, the elders carried out the sessions indoors, outdoors and sometimes between both indoors and outdoors. This movement, between the indoors and outdoors, impacted the learners’ visibility of the sessions. For example, elders would inadvertently move out of focus of the camera, as they assumed the iPad’s field of view was wide enough to display their entire work area. We also found that the participants’ experience of the session was influenced by the changing physicality of the environment. For example, in one of the sessions, where elders demonstrated how to make coconut oil, the indoor kitchen used was dark and smoky while the outdoor area was bright and sunny. This change in lighting and visibility influenced the learners’ experience of the session.

3. CONCLUSION
Despite use of mobile technologies in this study, the challenges experienced in managing the video sessions emphasised that mobile devices may not be best suited to mediate mobility across large indoor and outdoor spaces. One of the reasons for this is that mobile devices are designed to be carried around, while elders preferred not to handle devices while conducting the sessions. The inability of mobile devices to mediate the mobility of the participants in this context, thereby made the mobile devices immobile.

The consequence for design is therefore to use a mixture of both fixed and mobile devices that would allow hands free movement for the elders, adapt to the changing physicality of the environment and allow the remote learners to move with the elders through indoor and outdoor spaces.

Our future work involves building a prototype that meets the design requirements identified in this study, in order to support the practise of indigenous knowledge over distance.

4. REFERENCES
Gay men’s perceptions of mental health apps

[Extended Abstract]

Fernando Estrada Martinez de Alva
Department of Computing and Information Systems
The University of Melbourne
festrada@student.unimelb.edu.au

ABSTRACT
Mental health disorders are one of the most prevalent disabilities worldwide. Substantial research shows that at least 25% of the population may be currently struggling to cope with their mental health. Additionally, individuals who identify as gay men have a 30% higher risk of suffering from a mental health disorder. In this context, some gay men are using mobile apps to self-manage their mental health. Yet there is no adequate understanding of what type of mobile apps they use, and what problems they face and are trying to solve with the use of mobile apps. Likewise, we know little about gay men’s perceptions of the effectiveness of, and their satisfaction with, such apps.

In this qualitative study currently in progress we want to gain a detailed understanding of what type of apps gay men use, their reasons for using them and how they use them to self-manage their mental health. Additionally, we want to ascertain whether and how they recommend or engage with such apps and how they understand their own use of these apps. Through the findings of this study we aim to inform better design of apps for mental health that are inclusive of the gay community.

CCS Concepts
• Human-centered computing → Empirical studies in ubiquitous and mobile computing

Keywords
Mental health; Gay men; mobile apps

1. INTRODUCTION
Currently, the burden of mental health disorders leads the causes of disability in the world [1]. Mental disorders have been described as a person’s persistent inability to fully engage in work and study related tasks and social interactions [1]. Anxiety and depression are the most prevalent mental health disorders in which a person may experience fear of emotions, inability to modulate emotions and fear of social engagement.

Existing literature shows that when compared with heterosexuals, those who identify as gay men are twice as likely to experience anxiety and three times as likely to experience depression, and are less likely to use technology for mental health because of greater sensitivity to stigma [2]. This is an important problem, as people who identify as gay men are at higher risk of committing suicide [2].

2. APP USE BY GAY MEN
Some gay men are using mobile apps to self-manage their mental health [2]. Mobile apps such as mood diaries, mood tracking, mindfulness and even dating apps and networking apps like Facebook are being used by this group as a means to support their mental health [3]. Technologies for mental health have the potential to benefit people in this community by providing them with resources to connect socially, and access mental health services and information related to their recovery from mental health problems. However, in order to design technology-based services that are acceptable to this group, it is necessary to understand the problems and context (relevant to Mental Health) that gay men face.

3. RESEARCH DESIGN
We obtained ethics clearance from the University of Melbourne to conduct this qualitative study. The aim of this study is to gain an in-depth understanding of how gay men use mobile apps to self-manage their mental health (and not only apps designed specifically for mental health), and in general the problems they face and are trying to solve through the use of apps. Requirements to participate in this study are that participants self-identify as gay men, are over 18 years old of age and use or have used mobile apps to self-manage their mental health. We are conducting a 30 minute interview with each participant. We have interviewed six participants so far and aim to interview a total of 15. The purpose of the interview is to understand the participant’s background with apps for mental health, how the participant makes sense of such apps, and any non-use of such apps. Currently we are transcribing the interviews already undertaken. We are analysing data using a grounded theory approach.

4. TENTATIVE OUTCOMES
So far we have identified some commonalities in the use of apps by gay men. For instance, those whose personal values are related to nurturing relationships in real life prefer not to use dating apps as a means to engage in social interactions. Additionally, those who self-reflect on the purpose of using apps as the only means to support their mental health prefer to reach for direct mental health support.

5. REFERENCES
ABSTRACT

In this paper, we describe the rationale for, methodology and preliminary results of a survey designed to capture Melbourne children’s digital play habits.

CCS Concepts

• Human-centered computing → Empirical studies in collaborative and social computing

Keywords

Digital games; Children; Survey Methodology;

1. INTRODUCTION

Children’s use of screen based media (SBM) is a topic of debate in public and academic spheres [5][3]. Researchers from diverse fields seek to understand children’s relationship with SBM. One form of SBM that is receiving attention is digital gaming. Given the rapidity of technological and cultural change surrounding children’s media habits, there is a need for constantly updated information about what children do with their ‘screen-time’ [2]. This is needed to ensure that research about ‘effects’, for example, is relevant to contemporary childhoods.

2. BACKGROUND

Talk of the amount of time children spend with SBM significantly shapes how the debate is played out in public and academic spheres [2]. The specifics of what children actually do with screens are lost in existing device-level or genre-level data. Furthermore, digital gameplay, when considered largely as monolithic construct, is often assumed to be a solitary, sedentary and cognitively light use of time [1]. Scholars from diverse fields are calling for this assumption to be critically reassessed [4]. The popular game Minecraft makes for a timely case study and therefore, a survey to map the game’s use is discussed here.

3. METHODOLOGY

Parents of children aged between 3 and 12 years of age were invited to participate through schools, early learning centres, social media and word of mouth. A total of 752 participants completed the survey.

4. RESULTS

Preliminary descriptive statistics indicate that 90% of children in the sample as a whole play a wide range of digital games with titles like Minecraft, Lego, and Peppa Pig dominating. Tablet devices are preferred for digital play but PCs gain in popularity in the older age group (9-12yrs). 45% (N=305) of the sample had played Minecraft in the last month ‘a few times per week’ for ½ an hour to 1 hour on weekdays, and slightly longer on weekend days.

Further analysis will test for gender and age differences in Minecraft play in particular, as well as more in depth exploration of responses regarding other SBM use such as YouTube and television/movie format content.

5. REFERENCES


Tangible Sound-Based Enrichment for Orang-utans

[Extended Abstract]

Patricia Pons, Javier Jaen
Universitat Politècnica de València
ppons@dsic.upv.es, fjjaen@upv.es

Marcus Carter, Sarah Webber
Microsoft Research Centre for SocialNUI
{marcus.carter, s.webber}@unimelb.edu.au

ABSTRACT
This paper describes the design, development and upcoming evaluation of a tangible user interface for orang-utans. Grounding on primates’ abilities and preference towards the manipulation of objects, the developed system will augment the use of everyday non-technological elements with sound-based interactions. The behaviours of the orang-utans and explorations of the system will be studied in order to inform the design of future experiences for animal enrichment activities.

CCS Concepts
• Human-centred computing → User-centred design

Keywords
ACI, TUI, digital enrichment, interaction design

1. INTRODUCTION
The use of interactive technologies in zoos has opened a whole new area of possibilities to create enrichment activities. Their goal is to improve animals’ wellbeing while raising zoo visitors’ awareness. However, working with some species can be particularly challenging. Several restrictions have to be considered when developing technology at the zoo as it is essential to guarantee the welfare of the animals. In this regard, orang-utans strength, which by far exceeds that of humans, makes it extremely difficult to provide them with free access to technology inside their enclosure. As a result, the provided technology might not be as compelling and engaging as it could be, and the interaction may be less natural and limited to when the zoo keepers are present.

A pioneer project between the University of Melbourne and Melbourne Zoo, ‘Kinecting with Orang-utans’ [1], has recently studied how orang-utans interact with several applications consisting of digital projections on the floor of the enclosure. The technology here remains outside the enclosure to ensure the animals’ welfare. However, the interaction happens completely inside, allowing more freedom of interactions and explorations. It was observed that the animals eventually interacted with these projections using everyday objects they had around such as straw, blankets and plastic balls. Orang-utans have intrinsic abilities for the manipulation of objects and creation of tools, and objects and toys are also part of current zoo enrichment activities. Following this idea, this work proposes the use of non-technological everyday objects as mediators of the interactive experience. This could offer a wide new range of opportunities to explore digital enrichment at the zoo. Manipulations of objects are in this case augmented with sound-based stimuli, i.e., the movements of objects by the orang-utans generate different sounds according to the gestures performed, location and speed of the interactions.

2. DESIGN PROCESS AND STUDY
This research follows a user-centred approach with non-human animals as target users, as proposed by Mancini’s manifesto for Animal-Computer Interaction [2]. It is essential to involve the zoo keepers during the design and development process of these kinds of systems, as their suggestions and expertise provide very valuable knowledge to inform the design of the system. Due to the fact that the target users are unable to provide explicit and verbal feedback, behavioural observation and collaboration with zoo keepers and animal welfare specialists is necessary.

The developed system uses a Microsoft Kinect sensor to track the object being manipulated and the movements performed with it. In this way, the orang-uts can interact with the objects from any location and posture within the tracked area. The system is capable of tracking objects in the enclosure which can be distinguished by their colour. Hence, the interaction is not limited to a specific or instrumented object. Instead, this approach allows the orang-utans to explore the sound-based interactions with different toys or items they have around without requiring to put any special artefact on the object. After an iterative process of brainstorming and discussion with the zoo keepers, three different modalities for generating sound have been developed: a) continuous: different notes are generated continuously depending on the movements and distance to the sensor, b) progression: a song is played while the object keeps moving, and c) noise: noisy and clanking sounds are generated each time the object is moved. Volume variations of the sounds are also produced based on the speed of the movements performed. These three sound generation modalities will be tested in June and July 2016.

3. CONCLUSIONS AND FUTURE WORK
The evaluation of this system would allow to study how orang-utans interact and explore different sound enrichment activities and how do they use tangible objects to mediate the activity. The results of this study will inform the design of future multimodal experiences for orang-utans, such as using projections, sounds and tangibles together. The purpose of this enrichment activities could range from playful experiences to more challenging activities, such as matching games, to support the animals’ mental stimulation. In addition, interspecies activities between human visitors and orang-utans could be developed based on the common ability of object manipulation. This shared skill could help connecting visitors with the species, reinforcing zoos’ efforts to raise awareness about wildlife preservation and animal welfare.

4. ACKNOWLEDGMENTS
This work is funded by the EDRF-FEDER, Spanish MINECO (TIN2014-60077-R), Spanish MECD (FPU13/03831) and Australian Government (Endeavour Research Fellowship). The authors gratefully acknowledge the invaluable support of the Microsoft Research Centre for Social NUI and Zoos Victoria.

5. REFERENCES
Designing Avatars to Promote Social Engagement for Older Adults

[Extended Abstract]
Romina Carrasco Zuffi
Microsoft Research Centre for Social Natural User Interfaces, University of Melbourne
rcarrasco@student.unimelb.edu.au

ABSTRACT
Social engagement among older adults improves quality of life, reducing negative emotions that may lead to depression or premature death. The use of virtual avatars in online games can address this problem by providing opportunities for enjoyment, social engagement and self-representation. Avatars can affect the behavior of users in both the digital and the physical world. However, further study is needed to identify how to promote social engagement for older adults through appropriate design of avatars. This study aims to understand how older adults (65+) are using player representations in online games for socializing. The approach will firstly use qualitative methods to generate a holistic understanding of current practices of older adults using online games and secondly a participatory design approach to create configurable avatars.

CCS Concepts
Human-centered computing → Social networks; • Human-centered computing → Empirical studies in interaction design; • Human-centered computing → User studies;

Keywords
Older adults; online games; social inclusion; avatar; socializing

1. INTRODUCTION
The proportion of the older population is growing worldwide. Developed countries are already experiencing repercussions with low levels of social engagement among some older adults, which can be caused by mobility or sensory impairments, geographical isolation or family circumstances [1]. Social interactions can bring positive health consequences [1]. Thus it is becoming essential to find innovative ways to promote social involvement in later life. The increasing participation of older adults in online technologies may be an initial way of addressing this situation. In Australia, 39% of adults aged 65 to 94 are playing online games [2]. Participating in online games is one way in which digital technologies offer opportunities to enhance older adults’ social engagement.

Growing older can lead the body to become a ‘prison’ that does not allow individuals to maintain their lifestyles [3]. In order to confront this problem, researchers suggest that older adults could use digital domains to explore experiences that are limited by their physical bodies [3]. Their participation may be done through an “avatar” a visual animated representation of the person, which is controlled by the user in real time. This digital body has characteristics that can facilitate social behavior as it uses natural ways of expression as gestures, voice and body movements promoting bodily experiences to communicate with other people online without the requirement of meeting physically. Using the digital body as a way to perceive the world and experience activities that are constrained by the physical body may be a way of liberating older people’s minds and open possibilities of self-expression and social empowerment. It is also a way to override physical distances and geographical isolation that may be a limitation for others.

2. AIMS OF PROPOSED STUDY
The goal of this research is to identify the causes, context or circumstances that lead participants to create and use avatars for social engagement. This study aims to understand how representations of self in online social games provide older adults with opportunities to generate social interactions. In addition, it aims to understand the participants’ social behavior through online games, the perceived usefulness of their digital representation and their views on the benefits and consequences of the activity in their lives.

3. DESCRIPTION OF PROPOSED STUDY
The first study will use a qualitative exploratory approach based on semi-structured interviews and participant observations of playing behavior. The recruitment criteria will include older adults (65+) living in Melbourne who have been playing online games for at least three months with other people. Older people who speak English, Italian and Spanish will be eligible to participate and will be recruited via a combination of advertisement in online and offline communities and snowball sampling techniques. Only older people capable of giving informed consent will be recruited for the project.

The data collection will use a combination of audio and video recording to register the interview and the game playing behavior accordingly. Additionally, the researcher will make field notes during the whole session, photographs to register the context of playing and screenshots of the player’s online game. The data will be transcribed and anonymized and translated to English where necessary. The topics then would be classified and categorized using a thematic analysis relevant to the research question. These themes will be developed for publication and will also serve as the basis for further studies.

The study is waiting for ethical approval of Melbourne University to start the recruitment of participants.

4. REFERENCES
ABSTRACT

Drawing on knowledge-based theory of the firm that considers knowledge as the most strategically significant resource of the firm, this paper develops a theory on information security. It identifies the key concepts of information security, describes the relationships between these concepts, and explains how opposed tension between controls and threats either improves or degrades the usability of information-based resources. An outcome of creating a minimum viable resource is also discussed. The paper provides the theoretical base to understand why practitioners safeguard information in addition to implications for linking this theory to the theory on resource-based view of the firm.

CCS Concepts

- Information systems ➔ Security and privacy ➔ Formal methods and theory of security ➔ Formal security models

Keywords

Information security; resource based view; organisational security; controls; threats; competitive advantage; theory development

1. INTRODUCTION

The concept of information security is not new, however the reasons and motivations behind it are imperfectly understood. This manuscript seeks to explain how and why the phenomena that comprise the concepts of information security occur but does not attempt to create testable predictions. The emphasis for the rest of this paper is to comprehend the information security concepts and relationships between them in order to alter our understanding of why we continue with efforts to protect information.

In terms of a concise description, this proposed theory of information security simply states that the motivation behind all attempts by an organisation to secure information against threats is to create resources that can later be used for organisational performance. Information will degrade over time without adequate controls implemented for its protection. In terms of the taxonomy of IS theories presented by Gregor (2006), this manuscript provides a (Type 2) high-level theory for explanation, describing how and why the phenomenon of information security occurs. The main independent construct in this theory is information and the dependant construct is resources.

The theory of information security originates from the area of information systems, built entirely from concepts that relate to information and the breadth of systems that it can reside on. In terms of levels, it includes strategies to protect information used by individuals, groups, organisations and also protects information shared between organisations. This theory links to the resource-based view of the firm, supporting the potential use of information as a tangible resource, by an organisation for competitive advantage. Implications are that, depending on the information affected, degradation over time may reduce the usefulness of the resource and thus lead to the potential erosion of competitive advantage or organisational performance.

The paper continues in three major sections, with the structure adapted from Rivard (2014). Initially, we introduce information security, discuss why this theory of information security is needed and carefully examine the issues with existing theory. Second, we explain the theory of information security. Thirdly, we examine the implications for the development of this theory. Finally, we draw conclusions, consider limitations and offer proposals for future research to improve our understanding of information security.

2. INFORMATION SECURITY

This section describes why a new theory on information security is needed, based on an in-depth exploration of the philosophical and theoretical issues in existing theory. This develops the set of conditions that this new theoretical development satisfies.

3. THEORY ON INFORMATION SECURITY

This section describes the conceptual elements of the proposed theory, the relationships between the concepts, and implications for use of the theory.

The following schematic provides a visual description:

![Figure 1. Schematic of information security theory.](image)

4. REFERENCES


ABSTRACT
The process of information security risk management (ISRM) enables an organization to not only identify risks specific to its information and assets, but also to assess the impact and likelihood of a threat occurrence. Despite significant interest and investment in ISRM, its organizational practice still has deficiencies since it is not considered a standard management process of an organization and lacks evidence based decision making. Business analytics (BA) presents organizations with a unique opportunity to base their ISRM upon continuous monitoring and data analysis, thereby helping executives make timely, data driven security decisions in a proactive manner. In this research in progress paper, we utilize Resource -based Theory (RBT) to develop a research model that explains how interaction between the capabilities of security analytics and ISRM results in the development of an enhanced, high level analytics enabled ISRM. This in turn impacts overall security performance. We define the model based on an extensive analysis of the BA and ISRM literature. The model also provides a basis for future empirical work including focus groups, case studies and a survey.

CCS Concepts
• Information systems → Decision support systems

Keywords
Business Analytics; Risk Management; Capabilities; Security Performance; Information Security

1. INTRODUCTION
Organizations undertake information security risk management (ISRM) to identify and prioritize risks specific to their information assets along with assessing the impact and probability of threats occurring [1]. A review of information security literature reveals that ISRM practice still has deficiencies as it is not considered an integral part of key business processes and lacks evidence based decision making. This implies that business managers are not incorporating important security data into their decision making process [3], do not have holistic security awareness [4], and are therefore unable to make informed security related decisions.

Business Analytics (BA) deals with the information that is available for an enterprise to make informed decisions [2]. The capability to gain insights by analysing disparate data sources and very large volumes of both internal and external data is possible only with BA. However, despite strong anecdotal evidence; and optimistic reports of return on investment exceeding 100%, a systematic and structured analysis of employing BA for the enhancement of the ISRM process has not yet been conducted. We address this gap in current knowledge by asking the following research question:

How does the use of business analytics in information security risk management affect the overall performance of information security processes in an organisation?

To answer this question, we propose a model (see figure 1) that explains the impact of using BA (in ISRM) for an enhanced security performance.

Figure 1. Security performance impacts from analytics enabled ISRM capabilities

We extend the concept of BA to security analytics and examine how security analytics and ISRM capabilities interact with one another to develop an enhanced high level analytics enabled ISRM that contributes to superior security performance. This is important as information risks and security threats are increasing daily in terms of frequency and complexity, and organizations need to transition from reacting to incidents to anticipating threats and taking actions based on predictive analytics. Analytics enabled ISRM facilitates this transition with a potential to give a similar quality of actionable insights into information security as it does in marketing, supply chain management, customer relationship management and medical research [2].

2. REFERENCES
Some Things Change, Some Stay the Same: The Impact of Book Format on Borrowing Patterns

[Extended Abstract]

Dana McKay
Interaction Design Group, Computing and Information Systems, University of Melbourne
dmckay1@student.unimelb.edu.au

ABSTRACT
There has been considerable concern over how the introduction of ebooks impacts reading, but little on how it impacts book finding. This paper presents a simple study comparing print and ebook loan patterns within the same library, and shows that these finding interactions are demonstrably different.

CCS Concepts
•Human-centered computing~Human computer interaction (HCI)

Keywords
Ebooks; information seeking; HCI; libraries; browsing

1. INTRODUCTION
The work presented here is derived from [4], and examines differences in borrowing interactions between print and ebooks.

Ebooks have a number of perceived advantages over print books, particularly from the point of view of academic library users. They are available anytime and anywhere and (in theory at least) can be checked out by many users at the same time. Despite these advantages, though, a number of library users complain about ebooks, either due to reading interaction problems [5] or loss of the browsing experience at the shelves [2]. While it may seem that search can replicate the shelf experience, academic library users remain unconvinced [1].

Print books, conversely, can be checked out or missing when a user wants them and are physically cumbersome to carry around; they also require visitors to visit the library in person. These library visits, however, afford users the opportunity to browse—or forcibly expose them to subject-similar alternative books, depending on your viewpoint. While readers are frustrated by the physical limitations imposed by print books [1], they claim to like browsing [1; 2], and there is recent evidence in our own work that borrowing affects borrowing patterns [3].

What remained unclear until recently was whether book format actually affected borrowing patterns: does the convenience of access to ebooks influence their use, and is there any evidence to suggest that the browsing seen in print doesn’t happen online. This paper addresses those questions.

2. METHOD
This work extends our earlier log analysis method [3]. We compared contemporaneous borrowing patterns for print and ebooks in a single academic library. We examined the number of books borrowed; the days of the week on which books were borrowed; the number of books borrowed per user; and whether books tended to be borrowed by the same user in subject clustered groups.

3. RESULTS AND CONCLUSIONS
We discovered a number of differences between print and ebook borrowing. Print books were more likely to be borrowed on weekdays, ebooks on weekends. Print books were more likely to be borrowed in clusters of one or more books, ebooks were ‘borrowed’ alone. Print books borrowed together were closer together in subject classification than ebooks. Finally, Ebook users visited the library more frequently than print book users. In contrast, there was one significant similarity: the number of books borrowed per user remained largely the same between formats.

When examining the data closely, it is clear that there is a trade off between the convenience of ebooks and the richness of experience borrowing print books: Readers typically borrowed only one ebook at a time, but borrowed them all at hours, whereas print books were often borrowed in groups that would be co-located on the shelves. Users should not have to choose between convenience and richness, however, and this study demonstrates that search-only information access changes reader behaviour. How best to enrich online information seeking remains a topic for future work.

4. ACKNOWLEDGMENTS
Thanks to Swinburne University of Technology for allowing me to use their library data, and for their assistance in accessing it.

5. REFERENCES
ABSTRACT

Due to the recent increase of mobile device adoption in the workplace as part of knowledge-sharing activities, the risk of knowledge leakage risk (KLR) has also increased. This is especially the case for knowledge-intensive organizations operating in highly-competitive environments, where knowledge is the source of innovation and competitive advantage. Accordingly, organizations have a growing need to manage risk strategies in order to mitigate knowledge leakage incidents. This study aims to identify the determinants that influence KLR perceptions through the use of mobile devices and how such perceptions inform organizational KLR mitigation strategies to safeguard against such incidents. We take a contextual approach, drawing upon mobile device usage context literature, particularly “social context interaction framework” and “model of context in computer science” and organizing the constructs from human, organizational and technological perspectives to understand the contexts within which knowledge leakage occurs and propose a theoretical model that can aid organizations in developing such strategies.

CCS Concepts

• Security and privacy → Systems security → Human and societal aspects of security and privacy → Social aspects of security and privacy;

Keywords

Knowledge management, knowledge leakage, knowledge leakage risk, mobile context, mobile device.

1. INTRODUCTION

Due to the increasing adoption of boundary-spanning information technologies in the workplace, such as mobile devices, knowledge-intensive organizations operating in highly-competitive environments are presented with challenges of preventing increasing leakage of sensitive details such as intellectual property, trade secrets and business strategies (Ahmad et al, 2014). Recent literature shows how organizations are struggling with leakage through different avenues such as social media, cloud computing and portable data devices (Ahmad et al, 2015). Although much of the literature has focused on technical aspects of leakage (i.e., data and information), little research has been conducted on knowledge leakage through mobile devices (Agudelo et al, 2015). In knowledge-intensive organizations, where knowledge is the main source of innovation and competitive advantage, addressing knowledge leakage risk (KLR) becomes paramount as knowledge is considered a critical asset for sustained competitive advantage (Grant, 1996). However, despite its relevance to organizations, knowledge leakage has not been addressed in much of the knowledge management literature, as the focus has been traditionally on fostering knowledge sharing and creation of workflows within organizations (Bosua & Scheepers, 2007). This study addresses the gap found in the literature regarding KLR through mobile devices in knowledge-intensive organizations operating in highly-competitive environment by providing a model that assist with determining the factors that cause such risk posing the following question: _How does the perception of knowledge leakage risk through mobile devices inform organizations’ mitigation strategies?_ To answer this question, this study takes a context approach to understand how risk changes depending on the circumstances within which knowledge leakage occurs.

2. RESEARCH MODEL

![Figure 1. Proposed Research Model Adapted from Bradley & Dunlop (2005), Chen & Nath (2008) and Melville et al. (2004)](image)

3. REFERENCES


Individual Use of Enterprise 2.0

[Extended Abstract]

Diana Wong
University of Melbourne
dianaw@students.unimelb.edu

Rachelle Bosua
University of Melbourne
rachelle.bosua@unimelb.edu.au

Shanton Chang
University of Melbourne
shanton.chang@unimelb.edu.au

Sherah Kurnia
University of Melbourne
sherahk@unimelb.edu.au

ABSTRACT
The purpose of this paper is to delineate how individual uses Enterprise 2.0 and the impact on social capital dimensions. Often existing literatures outline use of Enterprise 2.0 through the lens of its intended benefits. These are good in framing a better understanding of the potential of Enterprise 2.0; however, looking at the individual use of Enterprise 2.0 from a theoretical lens will provide evident implications towards organisation benefits.

CCS Concepts
• Information systems $\rightarrow$ Information systems applications; social networking sites

Keywords
Enterprise social networking, Enterprise 2.0, social capital

1. INTRODUCTION
Enterprise social networking site, or more commonly known as Enterprise 2.0, refers to a private, or internal social networking site for business use. Individual use of Enterprise 2.0 can be described as the distinct use of the enterprise social networking platform based on individual’s attitudes and beliefs, social influence and self-efficacy (Bandura, 1977; Gu et al., 2013)

Use is a broad concept and can be considered from different perspectives. Richter and Koch (2008) identified 6 different social networking sites (SNS) use categories based on the functionalities - self-representation, contact management, context awareness, communicate, and network awareness. In a different study, Ellison et al. (2007) studied the intention to use social networking site (Facebook) and found that individuals are more interested in social browsing - to learn more about another individual they already know than to get to know new contacts.

Nonetheless, these researches represent the use of personal SNS, and are not well descriptive of the individual use of Enterprise 2.0. There are limited studies specific to the use of Enterprise 2.0. DiMicco et al. (2008) conducted a study on IBM’s Enterprise 2.0 and identified reasons for using Enterprise 2.0 - connect with their co-workers, career advancement, and campaigning for their projects. An extension to Richter and Koch (2008) research, Richter and Riemer (2009) studied the potential ways of using Enterprise 2.0 and identified 3 modes of use - identifying experts, building personal context, and fostering existing relationships. In Riemer and Richter (2012) research, the authors categorised the use of Enterprise 2.0 as socialising, organising, crowd sourcing, information sharing, and awareness creation. To understand the individual use of Enterprise 2.0, it is also important to distinguish the different types of users. Price and Shneiderman (2009) identified 4 distinct user types using the Reader-to-Leader Framework. The successive levels are reading, contributing, collaborating, and leading. Although their research is focused on social media (i.e. YouTube), it is highly relevant to provide a better understanding of online social participation. Therefore, this study adopts a qualitative case study research method, collecting data through semi-structured interviews (56), observations, and document analysis.

2. FINDINGS

3. REFERENCES
ABSTRACT

Supply chain (SC) is a network of organisations and activities that deal with the acquisition of raw material, production and distribution of final products to satisfy end user need [1]. Agility in Supply Chain is critical for organisations wanting to remain competitive in today’s dynamic business environment [3]. There is increasing interest in the use of Business Intelligence (BI) in the Supply Chain Management (SCM) context to improve SC Agility. However, there is limited research exploring BI contributions to SC Agility [2]. In this research-in-progress paper we propose a model based on a conceptual analysis of the literature showing how BI can help organisations achieve SC Agility by supporting the key SCM areas.

Based on two case studies so far, we have found that SCM processes use BI reports during SC decision making. SC plan involves all other SCM processes. BI report supports decision making during planning processes based on historical evidence and analysis. Sourcing team uses BI reports to analysis historical performance of supplier and KPI assessment. Product planning team uses BI reports to determine production volume based on demand forecast. BI reports are also used in delivery and customer return processes to analyse carrier and manufacturer performance respectively.

CCS Concepts
• Information systems → Business intelligence

Keywords
Supply Chain; Supply Chain Management; Supply Chain Agility; Business Intelligence; RBV; Agility; Business Intelligence Report

1. INTRODUCTION

Organizations are frequently forced to form a Supply Chain network to work collectively to meet customer demands in a dynamic market environment. The concept of Agile SC has been introduced to define the SC capability that enables an organization to respond to unpredictable changes and uncertainties in dynamic business environment [2]. Achieving SC Agility is challenging and has become a research topic of increasing interest over the last decade.

2. RESEARCH AIM

The aim of this study is to explore the role of BI in evidence based management of plan, source, make, deliver and return of SCM. Such BI supported evidence based SCM can result in better agility in SC [2]. Following conceptual research diagram has been proposed to guide the current study:

3. RESULTS (IN PROGRESS)

Case studies found strong evidence of use of BI tools in SC planning processes that eventually help organizations to achieve agility in SC. Global product planning teams of two case companies use BI reports to analyse historical sales volume, existing demand and return history to derive demand forecast in different markets. Such evidence based BI report helps organizations to reduce uncertainty in sourcing, production and delivery plan. Both case companies monitor Supplier KPI using BI reports based on historical performance trend which includes product fault, delivery time, meeting promised delivery date and quality.

Case studies provide evidence of highly increased responsiveness and competency in SC as a result of BI report use. While moderate influence on Flexibility and Quickness is offered by the use of BI reports in respective SCM processes.

4. REFERENCES

Understanding How Start-ups Gain a Competitive Advantage from Cloud Computing

[Extended Abstract]

Mohammed Alrokayan
Department of Computing and Information Systems,
The University of Melbourne
mohammed@alrokayan.com

Wally Smith
Department of Computing and Information Systems,
The University of Melbourne
wsmith@unimelb.edu.au

Sherah Kurnia
Department of Computing and Information Systems,
The University of Melbourne
sherahk@unimelb.edu.au

ABSTRACT
Small-and-medium enterprises (SMEs) play an important role in the national economies of most countries. However, there have been extremely high percentages of failure among SMEs around the globe, particularly in small businesses. In Australia, the Australian Bureau of Statistics indicates that on average 44 small businesses close permanently each day. There is no definite answer as to why we have this problem. However, according to many studies [1]–[3], lack of a competitive advantage is one of the major reasons.

CCS Concepts
• Information systems
• Social and professional topics

Keywords
Start-up, cloud computing, business model, competitive advantage

1. INTRODUCTION
Although there are several ways for firms to gain a competitive advantage, this study focuses on how cloud computing can help start-ups to gain a competitive advantage from a business model innovation perspective. The potential of business model innovation combined with the use of cloud computing have been successful in addressing many problems faced by SMEs. Cloud computing has unique characteristics like agility, affordability, accessibility, and scalability that might enable start-ups and entrepreneurs to gain a competitive advantage in the market.

Cloud computing has commenced recently with the availability of affordable technologies and easy access to network resources. We will study two types of cloud technologies: Software-as-a-Service (SaaS) and Platform-as-a-Service (PaaS)/Infrastructure-as-a-Service (IaaS) as shown in Figure 1.

Figure 1 shows our research framework as this research project aims to understand how cloud computing helps start-up firms gain competitive advantage. We seek to understand how start-up firms use cloud technology to innovate in the four components of the business model (value proposition, market, financial aspects and operations), which may lead to a competitive advantage.

Most of the academic studies on cloud computing are focused on the technical aspects of the cloud computing, while others are studying the financial impacts of the cloud computing on large organisations and governments but not on start-ups. The research gap is the lack of academic studies of how cloud computing helps start-ups gain competitive advantage. Therefore, the central research question is: What factors influence the use of cloud computing by start-up firms to achieve competitive advantage?

To address the above research questions, we have designed a conceptual framework (Figure 1) based on our extensive literature review in cloud computing, business model, business model innovation and competitive advantage.

2. REFERENCES
ABSTRACT
Organisations are not just operating in a highly competitive, dynamic and economically challenging market landscape, but also in a highly complex and sophisticated security threat landscape, that poses significantly heightened risks. Information security has consequently become increasingly relevant to the strategic outlook of organisations. In this research we seek to move beyond information security’s traditional inward looking paradigm of protection of information resources, and explore how organisations can achieve and/or sustain competitive advantage by developing an information security capability, embedded within the very fabric of the organisation. We will draw insights from the strategic management literature, particularly from the Resource Based View (RBV) and the dynamic capabilities perspective, to develop a strategic model of information security capability. We will then proceed to empirically test this model to demonstrate its contribution to the organisation’s overall performance.

CCS Concepts
- Social and professional topics → Management of computing and information systems
- Security and privacy → Formal methods and theory of security

Keywords
Information Security; Information Security Strategy; Information Security Capability; Resource Based View; Dynamic Capability; Competitive Advantage

1. INTRODUCTION
Organisations are not just operating in a highly competitive, dynamic and economically challenging market landscape, but also in a highly complex and sophisticated security threat landscape that exposes its information infrastructure to a range of security risks. Organisations may suffer reputational damage, loss of revenue, costs arising from breaches of confidentiality agreements and loss of productivity as a result of leakage of sensitive information [1]. Unsurprisingly, despite the existence of industry ‘best-practice’ security standards and unprecedented levels of investment in security infrastructure, the rate of incidents continues to escalate [2]. Thus significantly heightening the risks to organisations’ information resources.

Information security, as the function responsible for the protection of information by ensuring the confidentiality, integrity and availability of information assets, has consequently increased in relevance to become a business enabling, organisational level strategic function [3]. It has, however, maintained an inward looking focus, with no means of providing an opportunity for sustained competitive advantage to the organisation. However, with the increasing number, intensity, consequences and sophistication of information security attacks, information security now becomes critical to the survival of organisations. Organisations that desire not only to survive, but thrive, must recognise the value and role of the information security function. They must develop and articulate organisation-level information security strategies that integrate and embed information security within the fabric of the organisation. They must consciously develop and cultivate an organisational level information Security capability.

2. RESEARCH GAP
A review of security literature reveals an under developed strategic perspective of the Information Security Management function. Furthermore, there has been no reference to an organisational information security capability, and how this can be developed and utilised in achieving and/or sustaining competitive advantage.

3. RESEARCH AIM
This research looks to strategic management literature and attempts to draw insights from concepts such as the Resource Based View (RBV), and the dynamic capabilities perspective. It seeks to 1) define and describe an organisational Information Security capability, 2) to develop a strategic model that links resources with this capability, and then 3) empirically illustrate how it contributes to overall successful business performance.

4. REFERENCES
The Effect of Gaze on Gameplay in Co-located Multiplayer Gaming Environments

[Extended Abstract]

Joshua Newn
Microsoft Research Centre for Social NUI
The University of Melbourne
joshua.newn@unimelb.edu.au

ABSTRACT

Little is currently known about the effect of gaze in co-located multiplayer game environments such as tabletop board games. The configuration in which they are played allows for users to see one another face-to-face as well as their gaze direction. Opponents and collaborators alike can deduce where one another is looking, providing a form of non-verbal communication. Developments in low-cost eye tracking technology has allowed us to determine the gaze of users with acceptable accuracy. Recently, eye tracking technology has made it out of the laboratory and into the domestic settings but suffers from low rates of adoption. This research sets out to demonstrate the use of gaze in more social and playful situations. Users may very well change their strategies according to gaze information displayed on a shared space. In the same way, gaze can also be used as a form of deception. With the availability of eye tracking technology, researchers both in games and HCI alike are given the opportunity to examine the effects of multiuser gaze during gameplay, leading to future design considerations.

CCS Concepts

• Human-centered computing → Collaborative interaction

Keywords

Eye tracking; Gaze tracking; Games; Co-located play; Multiplayer

1. INTRODUCTION

The setting in which co-located multiplayer games are played such as tabletop board games encourages players to interact face-to-face with one another through collaboration, competition or cooperation. Speech and gestures are well-established modalities used for communication during gameplay, typically during one’s turn in a turn-based game. Gaze happens throughout the game, players monitor each other’s gaze during gameplay (Tse et al. 2007), providing a form of visual evidence towards attention, and subsequently insights into potential strategies and communication cues. Researchers have investigated the role of multiuser gaze behaviour, patterns and awareness when communicating around the tabletop (e.g. Vertegaal et al. 2001). However, there is a lack of investigation especially on multiplayer games and none on the effects of explicit gaze during gameplay.

The dynamics of gameplay may very well be influenced when gaze of players is made visible in different scenarios. Adaptive AI in computer games are able to adapt based on a player’s gaze (e.g. Wetzel et al. 2014) and it is possible for humans to adapt as well once this information is made visible. So far, gaze has been used to enhance gameplay through various mechanics but at the same time, suffers from low rates of adoption and is often seen as a gimmick. In order to change the way gaze is current perceived, it is essential that researchers think beyond gaze as an input method, but rather, as a form of revealing ones thought process and how it can be incorporated as a major component of the game.

2. RESEARCH DESIGN

To address the immediate research gap, I will first evaluate the how a multiplayer tabletop game will change when information based on gaze is made visible to opponents. Two players will be equipped with wearable eye trackers and play a multiplayer game (e.g. chess). The game board on the tabletop will be augmented with a top down projection and will display a variety of gaze visualisations. This will primarily consist of real-time heatmaps, scanpaths and coloured fixations on specific game pieces (where applicable). Rounds of no visualisation will be included. Each pair of players will play a total of three rounds to determine a winner. Recordings will be made throughout and will be used for analysis. Each player will be interviewed after the game and be asked to point out specific points in the game where gaze visualisation of the opponent has benefited them and when their gaze being visualised appeared to have disadvantaged them during the game. Players are expected to use their gaze as method of deception.

Alternatively, gaze can be used collaboratively, for example, by having an application which needs two users to gaze upon the same object. A simple application will be used to demonstrate this where the two users are asked to find item(s) in a dark room using a torch that is controlled by their gaze. Data will be collected to compare and contrasted, analysed and reported accordingly.

3. FUTURE WORK

The findings from the initial study will inform future studies, including the use of eye gaze in multiplayer immersive gaming environments such as in Augmented Reality (AR). This will open up a new environment that will allow users to interact face-to-face between the physical and the virtual, gaze in this scenario could very well benefit the interaction between two parties.

4. REFERENCES


People and Organisations

A Learner Centric Approach to a Knowledge Sharing Virtual Community for General Practitioners

[Extended Abstract]

Abdulaziz Murad
Department of Computing and Information Systems
University of Melbourne, Australia
amurad@student.unimelb.edu.au

ABSTRACT
As sources of medical information grow exponentially online, General Practitioners (GPs) have been seeking new ways of learning online for their medical education. Virtual communities have been developed that would let GPs interact with one another to share their experiences and knowledge regarding various topics. However, GPs are reluctant to use such means especially for their continuing medical education (CME). Currently having 50% participation is an accepted threshold for a successful virtual community of practice (VCoP). Hence, this research proposes deriving design principles from the literature for a knowledge sharing environment to increase participation and then use a learner centered design approach to design and build the learning platform. Evaluation of the principles will be done with a focus group of GPs and the expected outcome is a set of principles that has been validated by GP learners in a knowledge sharing environment.

CCS Concepts
•Human-centered computing → Social networking sites;
•Social and professional topics → Health information exchanges

Keywords
Virtual Community; General Practitioner; Medical Education; Knowledge Sharing

1. INTRODUCTION
General Practice has been recognized as an integral foundation for an efficient health system and knowledge sharing is critical to support GPs practice [1]. GPs recognize that online sources have the potential to help them with their CME if done with a group of professionals [4]. A group of GPs online can become a VCoP to share their knowledge and learn more about their practice [1]. However, GPs have shown to be unwilling to use VCoPs for their CME [4] and even when used participation falls as time passes making VCoPs unsustainable for a long period of time [1]. Essentially, a VCoP is based on social constructivism learning for the individuals involved [3]. Social constructivism learning states that learners need to develop an understanding of the work context and culture by being actively engaged [2]. Designing a VCoP for knowledge sharing should take social constructivism learning into consideration [3]. Therefore, this research will develop a set of design principles for a knowledge sharing VCoP for GPs’ CME with the fundamentals of social constructivism learning. The resulting concepts from the literature are grouped into themes and made into a conceptual model. A Learner Centered Design (LCD) approach has been taken to develop the Web based learning platform, an online discussion forum, and a prototype is being developed to observe GPs’ initial reactions.

3. DESIGN PRINCIPLES
One important factor to consider when designing the principles is that it is a VCoP, thus individual and group design principles must be separated. From the literature, two individual design principles were taken into consideration: Provide Profile Customization and Offer Easily Navigated and Searchable Platform. Four group design principles were recognized: Include a Diverse Range of Non-Competing Users; Deliver Facilitated CME Content; Equip Facilitators with Necessary Platform Tool; and Promote the Feeling of a Community.

4. FUTURE WORK
A focus group qualitative session will occur with the GPs to refine the design principles for the VCoP to support their CME knowledge sharing and learning. A change towards the design principles will likely occur from GPs’ feedback from the focus group sessions. A case study will then be used to validate the design principles and the LCD approach to use as a method to increase participation and sustain the VCoP for GPs’ CME.

5. ACKNOWLEDGEMENTS
My thanks to Professor John Wark for giving me the opportunity to join his project and investigate the knowledge sharing between GPs in a VCoP for their CME. In addition, my thanks to Associate Professor Dr.Reeva Lederman, Associate Professor Dr.Shanton Chang, and Dr.Rachelle Bosua for their supervision, support, and guidance.

6. REFERENCES
Thank You!

An enormous amount of planning goes into the organization of an event such as the CIS DC 2016. The CIS DC 2016 would not have been possible without the support and participation of graduate research students and staff in the Department of Computing & Information Systems.

We would especially like to acknowledge Justin Zobel, the Head of the Department of Computing & Information Systems, Rhonda Smithies and Julie Ireland for assisting the CIS DC 2016 organizing committee with the planning and co-ordination of this event.

Our dear supervisors who have pushed us to submit papers and posters for the colloquium and have supported us in our studies thus far.

Our sponsors, Australian Computer Society, Google, Microsoft, Victorian Life Sciences Computation Institute, Melbourne Networked Society Institute, the Microsoft Research Centre for Social Natural User Interfaces, the Melbourne School of Information and the Melbourne School of Engineering for their financial contributions to make the CIS DC 2016 possible.

Our keynote speaker Drs Patrick Maes, industry partners and colleagues from other institutions.

The panelists and judges at each session, who have volunteered their time to make this event a success.

A heartfelt thank you to all our graduate research students and staff who have participated in the CIS DC 2016.

Finally, we would like to acknowledge the CIS DC 2016 DC organizing committee for volunteering their time, skills and enthusiasm and also thank the 2015 committee for their support.

The 2016 CIS DC Planning Committee

Caesar Wu
Doris Hoogeveen
Mohammad Oloomi
Ibrahim Al-Mahdi
Elham Naghizadeh
A Special Thanks to Our Sponsors!
Download the program and proceedings from our website!


cis-dc@unimelb.edu.au