ExSS 2018: Workshop on Explainable Smart Systems

Brian Lim
Department of Computer Science, School of Computing
National University of Singapore
brianlim@comp.nus.edu.sg

Alison Smith
Decisive Analytics Corporation
Arlington, VA, USA
alison.smith@dac.us

Simone Stumpf
Centre for HCI Design, School of Mathematics, Computer Science and Engineering
City, University of London
Simone.Stumpf.1@city.ac.uk

ABSTRACT
Smart systems that apply complex reasoning to make decisions and plan behavior are often difficult for users to understand. While research to make systems more explainable and therefore more intelligible and transparent is gaining pace, there are numerous issues and problems regarding these systems that demand further attention. The goal of this workshop is to bring academia and industry together to address these issues. The workshop includes a keynote, poster panels, and group activities, towards developing concrete approaches to handling challenges related to the design, development, and evaluation of explainable smart systems.

Author Keywords
Explanations; visualizations; machine learning; intelligent systems; intelligibility; transparency.

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION
Smart systems that apply complex reasoning to make decisions and plan behaviour, such as clinical decision support systems, personalized recommendations, home automation, machine learning classifiers, robots and autonomous vehicles, are difficult for users to understand [1]. Textual explanations and graphical visualizations are often provided by a system to give users insight into what it is doing and why it is doing it [3,7,11,13]. Previous work has stressed the importance of explaining various aspects of the decision-making process to users [8], and these different kinds of intelligibility types – for example, *Confidence* [5,9] showing the probability of the diagnosis being correct, either as a percentage or a pie chart, and *Why* and *Why Not* [10] providing facts used in reasoning about the diagnosis – have been used in smart systems [6,10].

MOTIVATION, TOPICS AND CONTRIBUTION
Research to make smart systems explainable is gaining pace, partly stimulated through a recent DARPA call on Explainable AI (XAI) [2], which seeks to develop more explainable models and interfaces that allow users to understand, appropriately trust and interact with these new systems. However, there are numerous issues and problems regarding explainable smart systems that demand further attention, such as:

- What is an explanation? What should they look like?
- Are explanations always a good idea? Can explanations "hurt" the user experience, and in what circumstances?
- When are the optimal points at which explanations are needed for a particular system?
- How can we measure the value of explanations or how the explanation is provided? What human factors influence the value of explanations?
- What are "more explainable" models that still have good performance in terms of speed and accuracy?

This workshop brings together industry and academic researchers in the area of explainable smart systems to exchange perspectives, approaches, and results.

WORKSHOP OVERVIEW

Keynote Speaker
The workshop keynote will be provided by David Gunning. David Gunning is DARPA program manager in the Information Innovation Office (I2O) and manages the Explainable AI (XAI) [2] and the Communicating with Computers (CwC) programs. Prior to these programs, he managed the Personalized Assistant that Learns (PAL) project that produced Siri and the Command Post of the Future (CPoF) project that was adopted by the US Army as their Command and Control system for use in Iraq and Afghanistan. He has previously worked at Pacific Research National Lab (PNNL), the Palo Alto Research Center (PARC), Vulcan Inc. and the Air Force Research Labs.

Accepted Papers
Fifteen papers were accepted to ExSS 2018 after a peer-review process~~:~~; each paper was reviewed by three members of the Program Committee:

- Enrico Bertini, —New York University, USA
- Maya Cakmak, University of Washington, USA
- Fan Du—, University of Maryland, USA
- Dave Gunning—, DARPA, USA
- Judy Kay—, University of Sydney, Australia
- Bran Knowles—, University of Lancaster, UK
- Todd Kulesza—, Microsoft, USA
- Mark W. Newman—, University of Michigan, USA
- Deokgun Park—, University of Maryland, USA

- Forough Poursabzi-Sangdeh—, University of Colorado, Boulder, USA
- Jo Vermeulen—, Aarhus University, Denmark

The papers will be presented during the themed poster panel session, which is organized into five themes:[1]

- *How to glean explainable information from machine learning systems* – "The design and validation of an intuitive confidence measure" (van der Waa et al.), "An Axiomatic Approach to Linear Explanations in Data Classification" (Sliwinski et al.), "Explaining Contrasting Categories" (Pazzani et al.), Explaining Complex Scheduling Decisions" (Ludwig et al.).
- *Explainable/semantically meaningful features* – "Explainable Movie Recommendation Systems by using Story-based Similarity" (Lee and Jung), "Labeling images by interpretation from Natural Viewing" (Guo et al.)
- *How to design and present explanations* – "Normative vs. Pragmatic: Two Perspectives on the Design of Explanations in Intelligent Systems" (Eiband et al.), "Explaining Recommendations by Means of User Reviews" (Donkers et al.), "What Should Be in an XAI Explanation? What IFT Reveals" (Dodge et al.), "Interpreting Intelligibility under Uncertain Data Imputation" (Lim et al.)
- *Explanations' impact on user behavior and experience* – "Explanation to Avert Surprise" (Gervasio et al.), "Representing Repairs in Configuration Interfaces: A Look at Industrial Practices" (Leclercq et al.), "Explaining smart heating systems to discourage fiddling with optimized behavior" (Stumpf et al.)
- *User feedback/interactive explanations* – "Working with Beliefs: AI Transparency in the Enterprise" (Chander et al.), "The Problem of Explanations without user Feedback" (Smith and Nolan)

The main part of the workshop is structured around two hands-on activity sessions in small subgroups of 3-5 participants. The activities are grounded in example systems provided by industry participants. The first session identifies challenges and high-level approaches for the example systems while the second session in explores concrete explanation or study designs for the example systems. Each of the subgroups works on the activities in parallel, and the outcomes are shared in a final presentation session.

Workshop Organizers

Dr. Brian Lim is an Assistant Professor in the Department of Computer Science at the National University of Singapore (NUS), Singapore, where he researches ubiquitous computing and intelligible data analytics for healthcare and smart cities [8–10]. He is also Principal Investigator at both the Biomedical Institute for Global Health Research & Technology (BIGHEART) and the Sensor-enhanced Social Media Centre (SeSaMe) at NUS.

Alison Smith is the Lead Engineer of the Machine Learning Visualization Lab for Decisive Analytics Corporation, where her focus is on enhancing end users' understanding and analysis of complex data without requiring expertise in data science or machine learning. She is also a PhD student at the University of Maryland, College Park, and her research focuses on human-centred design for interactive machine learning [12].

Dr. Simone Stumpf is a Senior Lecturer (Associate Professor) at City, University of London, UK, where she researches designing end-user interactions with intelligent systems [4,6,14]. Her current projects include designing user interfaces for smart heating systems and smart home self-care systems for people with dementia or Parkinson's disease.

REFERENCES

1. Alyssa Glass, Deborah L. McGuinness, and Michael Wolverton. 2008. Toward establishing trust in adaptive agents. In *Proceedings of the 13th international conference on Intelligent user interfaces - IUI '08*, 227. https://doi.org/10.1145/1378773.1378804

2. Dave Gunning. 2016. Explainable Artificial Intelligence (XAI). Retrieved December 20, 2017 from https://www.darpa.mil/program/explainable-artificial-intelligence

3. Jonathan L. Herlocker, Joseph A. Konstan, and John Riedl. 2000. Explaining collaborative filtering recommendations. In *Proceedings of the 2000 ACM conference on Computer supported cooperative work - CSCW '00*, 241–250. https://doi.org/10.1145/358916.358995

4. Todd Kulesza, Margaret Burnett, Weng-Keen Wong, and Simone Stumpf. 2015. Principles of Explanatory Debugging to Personalize Interactive Machine Learning. In *Proceedings of the 20th International Conference on Intelligent User Interfaces - IUI '15*, 126–137. https://doi.org/10.1145/2678025.2701399

5. Todd Kulesza, Simone Stumpf, Margaret Burnett, Weng Keen Wong, Yann Riche, Travis Moore, Ian Oberst, Amber Shinsel, and Kevin McIntosh. 2010. Explanatory debugging: Supporting end-user debugging of machine-learned programs. In *Proceedings - 2010 IEEE Symposium on Visual Languages and Human-Centric Computing, VL/HCC 2010*, 41–48. https://doi.org/10.1109/VLHCC.2010.15

6. Todd Kulesza, Simone Stumpf, Weng-Keen Wong, Margaret M. Burnett, Stephen Perona, Andrew Ko, and Ian Oberst. 2011. Why-oriented end-user debugging of naive Bayes text classification. *ACM Transactions on Interactive Intelligent Systems* 1, 1: 1–31.

[1] The papers are also published in this order.

https://doi.org/10.1145/2030365.2030367

7. Carmen Lacave and Francisco J. Díez. 2002. A review of explanation methods for Bayesian networks. *Knowledge Engineering Review 17*, 107–127. https://doi.org/10.1017/S026988890200019X

8. Brian Y. Lim and Anind K. Dey. 2010. Toolkit to support intelligibility in context-aware applications. In *Proceedings of the 12th ACM international conference on Ubiquitous computing - Ubicomp '10*, 13. https://doi.org/10.1145/1864349.1864353

9. Brian Y. Lim and Anind K. Dey. 2011. Investigating intelligibility for uncertain context-aware applications. In *Proceedings of the 13th international conference on Ubiquitous computing - UbiComp '11*, 415. https://doi.org/10.1145/2030112.2030168

10. Brian Y. Lim, Anind K. Dey, and Daniel Avrahami. 2009. Why and why not explanations improve the intelligibility of context-aware intelligent systems. *Proceedings of the 27th international conference on Human factors in computing systems - CHI 09*: 2119–2129. https://doi.org/10.1145/1518701.1519023

11. Pearl Pu and Li Chen. 2006. Trust building with explanation interfaces. In *Proceedings of the 11th international conference on Intelligent user interfaces - IUI '06*, 93. https://doi.org/10.1145/1111449.1111475

12. Alison Smith, Varun Kumar, Jordan Boyd-Graber, Kevin Seppi, and Leah Findlater. 2018. Closing the Loop: User-Centered Design and Evaluation of a Human-in-the-Loop Topic Modeling System. In *Intelligent User Interfaces*.

13. William Swartout, Cecile Paris, and Johanna Moore. 1991. Explanations in knowledge systems: Design for Explainable Expert Systems. *IEEE Expert* 6, 3: 58–64. https://doi.org/10.1109/64.87686

14. K Yarrow and I Sverdrup-Stueland. 2004. Fixing the Program My Computer Learned: Barriers for End Users, Barriers for the Machine Todd. *Openaccess.City.Ac.Uk* 47, May: 552–567. https://doi.org/10.1007/978-3-540-25939-8