

SUM'20: State-based User Modelling

Sahan Bulathwela
m.bulathwela@ucl.ac.uk
University College London
London, UK

Davor Orlic
davor.orlic@ijs.si
Knowledge 4 All Foundation
London, UK

María Pérez-Ortiz
maria.perez@ucl.ac.uk
University College London
London, UK

Colin de la Higuera
cdlh@univ-nantes.fr
Nantes University
Nantes, France

Rishabh Mehrotra
rishabh@spotify.com
Spotify Research
London, UK

John Shawe-Taylor
j.shawe-taylor@ucl.ac.uk
University College London
London, UK

Emine Yilmaz
emine.yilmaz@ucl.ac.uk
University College London
London, UK

CCS CONCEPTS

• **Information systems** → **Personalization; Task models; Recommender systems.**

KEYWORDS

Task-based search; state-aware systems; human in the loop; recommender systems; personalisation; user modelling.

Workshop Website: <https://www.k4all.org/event/wsdsmsum20>

1 MOTIVATION

Capturing and effectively utilising user states and goals is becoming a timely challenge for successfully leveraging intelligent and user-centric systems in different web search and data mining applications. Examples of such systems are conversational agents, intelligent assistants, educational and contextual information retrieval systems, recommender/match-making systems and advertising systems, all of which rely on identifying the user state in order to provide the most relevant information and assist users in achieving their goals.

There has been, however, limited work towards building such state-aware intelligent learning mechanisms. Hence, devising information systems that can keep track of the user's state has been listed as one of the grand challenges to be tackled in the next few years [1]. It is thus timely to organize a workshop that re-visits the problem of designing and evaluating state-aware and user-centric systems, ensuring that the community (spanning academic and industrial backgrounds) works together to tackle these challenges.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

WSDM '20, February 3–7, 2020, Houston, TX, USA

© 2020 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-6822-3/20/02.

<https://doi.org/10.1145/3336191.3371883>

2 SCOPE & OBJECTIVES

This workshop aims to facilitate a forum for leveraging intelligent methods for modelling user's state and its transition in an effective manner. The aim is to focus on several open questions:

- (i) **User representation and recommender systems:** How to design accurate state-aware representations of the user and embed that information in predictive models?
- (ii) **Task understanding and supporting user tasks:** How to identify the specific sub-task a user is trying to accomplish and design state-aware intelligent systems that assist users in each step of their goals?
- (iii) **State-aware evaluation:** How to develop metrics and evaluation techniques that model and understand user states, so as to provide more sensitive and enhanced evaluation of the user-centric systems?
- (iv) **Human in the loop:** How to include the human in the learning "loop" in an interactive fashion, making feedback an integral part of the modelling?
- (v) **User-aware systems:** How to incorporate user state information to achieve better human-machine intelligence?
- (vi) **Cognitive/contextual user understanding:** How to understand from a cognitive point of view how actions taken by users and their context can influence their state?
- (vii) **State-aware machine learning:** How do the different machine learning algorithms that enable to identify and leverage user states compare for different application scenarios?

To answer these questions, it is imperative for research communities from cognitive science, intelligent tutoring systems, information retrieval, human computer interaction, psychology and other diverse fields to bring forward their inputs, shaping the future of many personalised web search and data mining applications.

3 EXPECTED OUTCOMES

Rather than focusing on different aspects of state-based user modelling in isolation (such as user interactions, task taxonomy spaces, etc.), SUM'20 encourages tabling the challenges, findings and solutions relating to all the different aspects in one forum where all attendees are active participants. Given the focus areas of the

workshop, participants from academic institutions, industrial research labs as well as start-ups would benefit from this workshop. Specifically, the workshop will feature a mix of contributed talks, a panel discussion and invited talks by leading researchers from diverse backgrounds working in this theme. A specific segment of the schedule is reserved for the presentation of open problems followed by open discussions explicitly aimed at sparking collaborations between attendees to address these issues.

The workshop website is used as the communication channel to publish all relevant information, including accepted contributions and a summary of the discussion and findings from the workshop.

4 RELATED WORKSHOPS

There have been several workshops that can be considered of relevance to the SUM'20 research themes. These comprise: Task Intelligence Workshop¹ (WSDM'19), LearnIR: Workshop on Learning from User Interactions² (WSDM'18), several workshops at the Conf. on User Modeling, Adaptation and Personalization³ (UMAP'19), Where is the Human? Bridging the Gap Between AI and HCI Workshop⁴ (CHI'19), Interactive Machine Learning and Semantic Information Retrieval⁵ (ICML'17), Overview of the TREC Tasks Track 2016⁶ (TREC'16), The Future of Interactive Machine Learning⁷ (NIPS'16), ML for Education⁸ (NIPS'16), Human Centered Machine Learning⁹ (CHI'16) and Intelligent Techniques for Web Personalization and Recommendation¹⁰ (AAAI'14). These very recent workshops and discussion forums demonstrate that this is an opportune time to host a workshop focused on user-state modelling, and that this topic is recognised as a pressing one by many leading researchers in conferences related to web search and data mining.

5 BIOGRAPHIES OF ORGANIZERS

Sahan Bulathwela is a PhD. student at University College London (UCL) focussing on the theme: "Improving Recommendations of Educational Contents to Lifelong Learners" and currently contributes to the X5GON project (www.x5gon.org). Sahan's current research is geared towards modelling the knowledge state of lifelong learners using their learning journey and how these learner models can be used to identify the most helpful educational resources for them.

María Pérez-Ortiz is a Research Fellow at University College London (UCL). Prior to that she was Research Associate at the University of Cambridge (UK), a lecturer at University Loyola Andalucía and a Research Assistant at the Spanish National Research Council and the University of Cordoba (all three in Spain). María's PhD dissertation has received several awards (among others from the Spanish Scientific Society for Computer Science and the Spanish Association for Artificial Intelligence).

Rishabh Mehrotra is a Senior Research Scientist at Spotify Research in London focusing on ML problems for recommendations

in multi-stakeholder marketplaces. He obtained his PhD in the field of Machine Learning and Information Retrieval from University College London where he was partially supported by a Google Faculty Research Award. Some of his recent work has been published at top conferences including WWW, SIGIR, NAACL, CIKM, RecSys and WSDM. He is also a co-coordinator of the Tasks Track in TREC 2015, 2016 and 2017 and has co-taught a number of tutorials at leading conferences (including CIKM WWW).

Davor Orlic an MA graduate in Digital Humanities from University College London, co-founded VideoLectures.Net with 25.000 educational videos, created OpeningupSlovenia national educational test bed, established the UNESCO Chair in Open Technologies for OER and Open Education, and UNESCO Chair in Artificial Intelligence and now managing Knowledge 4 All Foundation with 60 global members. Works at the intersection of research, technologies, policies and research innovation in Artificial Intelligence.

Colin de la Higuera is Professor and UNESCO Chairholder in Technologies for the Training of Teachers by Open Educational Resources at University of Nantes. He has been involved in a number of research themes, including algorithmics, formal language theory and pattern recognition. He has been chairman of the International Community in Grammatical Inference (2002-2007), the founding president of the SIF, the French Informatics Society and is currently a trustee of the Knowledge 4 All Foundation and is working towards the usage of technology for an open dissemination of knowledge.

John Shawe-Taylor has contributed to fields ranging from graph theory through cryptography to statistical learning theory and its applications. However, his main contributions have been in the development of the analysis and subsequent algorithmic definition of principled machine learning algorithms founded in statistical learning theory, driving the mapping of these approaches onto novel domains including work in computer vision, document classification, and applications in biology and medicine focused on brain scan, immunity and proteome analysis. He published over 300 papers and two books that attracted over 60000 citations.

Emine Yilmaz is a Turing Fellow and Professor at University College London (UCL), Department of Computer Science, as well as an Amazon Scholar at Amazon Cambridge. Between 2012 and 2019, she also worked as a research consultant for Microsoft Research Cambridge, where she used to work as a full time researcher prior to joining UCL. Her research until now has received several awards including a Bloomberg Data Science Research Award in 2018, the Karen Sparck Jones Award in 2015 and the Google Faculty Research Award in 2014. Emine's current research interests include information retrieval, data mining and applications of machine learning. She has published research papers extensively at major venues such as ACM SIGIR, CIKM and WSDM, given several tutorials as part of top conferences, and organised various workshops.

6 ACKNOWLEDGEMENTS

This research is conducted as part of the X5GON project funded from the EU's Horizon 2020 research programme grant No 761758 and partially funded by the EPSRC Fellowship titled "Task Based Information Retrieval", under grant No EP/P024289/1. This material is based upon work supported by, or in part by, the U.S. Army Research Laboratory and the U. S. Army Research Office, and by the U.K.

¹<https://dl.acm.org/citation.cfm?doid=3289600.3291374>

²<http://rishabhmehrotra.com/WSDM18-learnIR.pdf>

³<http://www.cyprusconferences.org/umap2019/pages/workshops.html>

⁴<https://michae.lv/ai-hci-workshop/>

⁵<https://icml.cc/Conferences/2017/Schedule?showEvent=6>

⁶<http://trec.nist.gov/pubs/trec25/papers/Overview-T.pdf>

⁷<https://neurips.cc/Conferences/2016/Schedule?showEvent=6224>

⁸<https://neurips.cc/Conferences/2016/Schedule?showEvent=6241>

⁹<http://hcm12016.goldsmithsdigital.com/program/>

¹⁰<https://www.aaai.org/Press/Reports/Workshops/ws-13-11.php>

Ministry of Defence and the U.K. Engineering and Physical Sciences
Research Council (EPSRC) under grant number EP/R013616/1.

REFERENCES

- [1] J Shane Culpepper, Fernando Diaz, and Mark D Smucker. 2018. Research frontiers in information retrieval: Report from the third strategic workshop on information retrieval in lorne (swirl 2018). In *ACM SIGIR Forum*, Vol. 52. ACM, 34–90.