ALTA 2020

Proceedings of the 18th Workshop of the Australasian Language Technology Association

14–15 January, 2021
Virtual Workshop

Sponsors

Platinum

Australian Government
Department of Defence

Silver

SINTELIX

Google

Introduction

Welcome to the 18th edition of the Annual Workshop of the Australasian Language Technology Association (ALTA 2020). The purpose of ALTA is to promote language technology research and development in Australia and New Zealand. Every year ALTA hosts a workshop which is the key local forum for disseminating research in Natural Language Processing and Computational Linguistics, with presentations and posters from students, industry, and academic researchers. This year ALTA is hosted as a virtual workshop, due to the COVID-19 pandemic.

In total we received 25 paper submissions and we accepted 8 long papers and 8 short papers to appear in the workshop, as well as 3 extended abstracts. Of all submissions, 19 were first-authored by students. We had submissions from a total of five countries: Australia, New Zealand, Japan, Sri Lanka and United States. We are extremely grateful to the Programme Committee members for their time and their detailed and helpful comments and reviews. This year we had committee members from all over the globe including Australia, New Zealand, Japan, Sweden, Switzerland, United States and United Arab Emirates.

Overall, there will be two oral presentation sessions and two virtual poster sessions. We also ran a shared task in detection of human behaviour organised by Diego Mollá-Aliod (University of Macquarie), In addition, for the first time ALTA will have a Doctoral Consortium, with a single session for final year PhD students and recent PhD graduates to present their work to help young researchers to gain visibility to their prospective employers. Finally, the workshop will feature keynotes from Kendra Vant (Xero) and Andrew Perfors (University of Melbourne), following a tradition of bringing speakers from both academia and industry.

ALTA 2020 is very grateful for the financial support generously offered by our sponsors. Without their contribution, the running of these events to bring together the NLP community of the Australasian region would have been a challenge. We would like to express sincere gratitude to our sponsors.

We very much hope that you will have an enjoyable and inspiring time at ALTA 2020!

Maria Kim and Daniel Beck

Organisers:

Program Co-Chair: Maria Kim, Defence Science and Technology Group
Program Co-Chair: Daniel Beck, The University of Melbourne
Program Advisor: Meladel Mistica, The University of Melbourne

Program Committee:

Timothy Baldwin, Jennifer Biggs, Benjamin Boerschinger, Wray Buntine, Lawrence Cavedon, Trevor Cohn, Xiang Dai, Lea Frermann, Gholamreza Haffari, Hamed Hassanzadeh, Brian Hur, Nitin Indurkhya, Antonio Jimeno-Yepes, Sarvnaz Karimi, Sunghwan Mac Kim, Alistair Knott, Yitong Li, David Martinez, Nitika Mathur, Diego Mollá-Aliod, Scott Nowson, Cecile Paris, Lizhen Qu, Afshin Rahimi, Abeed Sarker, Andrea Schalley, Rolf Schwitter, Laurianne Sitbon, Kristin Stock, Hanna Suominen, Karin Verspoor, Stephen Wan, Michael Witbrock, Hiyori Yoshikawa, Xiuzhen Zhang

Invited Speakers:

Kendra Vant, Xero
Andrew Perfors, The University of Melbourne

Invited Talks

Kendra Vant (Xero)

Commercial machine learning at scale - the joys and the pitfalls

The art and science of applying machine learning techniques inside a for profit company is a world away from pursuing algorithm improvement and fundamental in a research setting. I will talk about the end to end process of building smart products within a SaaS company today.

Andrew Perfors (The University of Melbourne)

Beyond corpus data: Language as the result of active, theory-driven, environmentally-grounded inference

Most NLP approaches use external language resources, such as text corpora, to derive the distributional properties of word usage and represent linguistic meaning. In this talk I will review work from cognitive science exploring to what extent linguistic meaning depends on other factors as well, and how to capture them computationally. In the first part of the talk I will compare standard word-embedding models derived from corpus data to a semantic network derived from an extensive dataset of word associations involving more than 12,000 cue words and over 500K participants. I'll demonstrate that the word embedding model fails to capture important aspects of people's lexical representations that are captured by the word-association-based semantic network – aspects which probably reflect environmentally-grounded sensory knowledge as well as pragmatic and emotional understanding. In the second half, I will review evidence suggesting that human language learning involves active exploration and sophisticated conceptual/social reasoning in addition to bottom-up distributional mechanisms. Implications for NLP and computational linguistics will be discussed.

PROGRAMME - All times are in AEDT (Melbourne/Sydney).

14th December (Thursday) - Day 1

10:30 - 11:00 Opening Remakrs

11:00 - 12:00 Keynote: Kendra Vant (Xero)
Commercial machine learning at scale - the joys and the pitfalls

12:00 - 13:00 Lunch & Virtual Networking

13:00 - 14:00 Session 1 – Long Papers 1 (Session Chair: Cécile Paris)

Domain Adaptative Causality Encoder
Farhad Moghimifar, Gholamreza Haffari and Mahsa Baktashmotlagh
Automated Detection of Cyberbullying Against Women and Immigrants and Cross-domain Adaptability
Thushari Atapattu, Mahen Herath, Georgia Zhang and Katrina Falkner
The Influence of Background Data Size on the Performance of a Score-based Likelihood Ratio System: A Case of Forensic Text Comparison
Shunichi Ishihara
Feature-Based Forensic Text Comparison Using a Poisson Model for Likelihood Ratio Estimation
Michael Carne and Shunichi Ishihara

14:00 - 15:30 Afternoon Break & Poster Session (includes papers from Session 1)

Learning Causal Bayesian Networks from Text
Farhad Moghimifar, Afshin Rahimi, Mahsa Baktashmotlagh and Xue Li
Information Extraction from Legal Documents: A Study in the Context of Common Law Court Judgements
Meladel Mistica, Geordie Z. Zhang, Hui Chia, Kabir Manandhar Shrestha, Rohit Kumar Gupta, Saket Khandelwal, Jeannie Paterson, Timothy Baldwin and Daniel Beck
Benchmarking of Transformer-Based Pre-Trained Models on Social Media Text Classification Datasets
Yuting Guo, Xiangjue Dong, Mohammed Ali Al-Garadi, Abeed Sarker, Cecile Paris and Diego Mollá Aliod
Pandemic Literature Search: Finding Information on COVID-19
Vincent Nguyen, Maciek Rybinski, Sarvnaz Karimi and Zhenchang Xing
Leveraging Discourse Rewards for Document-Level Neural Machine Translation
Inigo Jauregi Unanue, Nazanin Esmaili, Gholamreza Haffari and Massimo Piccardi
The Open Domain Interviewing Agent

Ming-Bin Chen and Michael Witbrock

Cost-effective Selection of Pretraining Data: A Case Study of Pretraining BERT on Social Media

Xiang Dai, Sarvnaz Karimi, Ben Hachey and Cécile Paris

15:30 - 16:15 Session 2 – Doctoral Consortium (Session Chair: Stephen Wan)

Recognizing Biomedical Names: Challenges and Solutions

Xiang Dai

Ngana Wubulku Junkurr-Jiku Balkaway-Ka: The Intergenerational Co-Design of a Tangible Technology to Keep Active Use of the Kuku Yalanji Aboriginal Language Strong

Jennyfer Lawrence Taylor

Automatic Generation of Security-Centric Description for Cyber Threats

Tingmin Wu

16:15 End of ALTA 2020 Day 1

15th January (Friday) Day 2

11:00 - 12:00 Keynote: Andrew Perfors (The University of Melbourne)

Beyond corpus data: Language as the result of active, theory-driven, environmentally-grounded inference

12:00 - 13:00 Lunch & Virtual Networking

13:00 - 14:00 Session 3 – Long Papers 2 (Session Chair: Trevor Cohn)

Modelling Verbal Morphology in Nen

Saliha Muradoglu, Nicholas Evans and Ekaterina Vylomova

An Automatic Vowel Space Generator for Language Learner Pronunciation Acquisition and Correction

Xinyuan Chao, Charbel El-Khaissi, Nicholas Kuo, Priscilla Kan John and Hanna Suominen

ABSA-Bench: Towards the Unified Evaluation of Aspect-based Sentiment Analysis Research

Abhishek Das and Wei Emma Zhang

A machine-learning based model to identify PhD-level skills in job ads

Li'An Chen, Inger Mewburn and Hanna Suonimen

14:00 - 15:30 Afternoon Break & Poster Session (includes papers from Sessions 2 & 3)

Transformer Semantic Parsing

Gabriela Ferraro and Hanna Suominen

Convolutional and Recurrent Neural Networks for Spoken Emotion Recognition

Aaron Keesing, Ian Watson and Michael Witbrock

Popularity Prediction of Online Petitions using a Multimodal DeepRegression Model

Kotaro Kitayama, Shivashankar Subramanian and Timothy Baldwin

Exploring Looping Effects in RNN-based Architectures

Andrei Shcherbakov, Saliha Muradoglu and Ekaterina Vylomova

15:30 - 17:00 Session 4 – Shared Task, AGM, Best Papers and Closing

Shared Task (Chair: Diego Mollá-Aliod)
AGM (Chair: Sarvnaz Karimi)
Best Paper Awards (Chair: Maria Kim)
Closing (Chair: Daniel Beck)

17:00 End of ALTA 2020 Day 2

Table of Contents

Long Papers

Short Papers

Shared Task (Not Peer Reviewed)