ACL 2018

Deep Learning Approaches for Low-Resource Natural Language Processing (DeepLo)

Proceedings of the Workshop

July 19, 2018
Melbourne, Australia

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 978-1-948087-47-6

Preface

The ACL 2018 Workshop on Deep Learning Approaches for Low-Resource Natural Language Porcessing took place on Thursday July 19, in Melbourne Australia, immediately following the main conference.

Natural Language Processing is being revolutionized by deep learning with neural networks. However, deep learning requires large amounts of annotated data, and its advantage over traditional statistical methods typically diminishes when such data is not available; for example, SMT continues to outperform NMT in many bilingually resource-poor scenarios. Large amounts of annotated data do not exist for many low-resource languages, and for high-resource languages it can be difficult to find linguistically annotated data of sufficient size and quality to allow neural methods to excel. Our workshop aimed to bring together researchers from the NLP and ML communities who work on learning with neural methods when there is not enough data for those methods to succeed out-of-the-box. Techniques of interest include self-training, paired training, distant supervision, semi-supervised and transfer learning, and human-in-the-loop algorithms such as active learning.

Our call for papers for this inaugural workshop met with a strong response. We received 22 paper submissions, of which 6 were “extended abstracts”—work that will be presented at the workshop, but will not appear in the proceedings in order to allow it to be published elsewhere. We accepted 10 papers and 5 extended abstracts. Our program covers a broad spectrum of applications and techniques. It was augmented by invited talks from Trevor Cohn (Melbourne), Sujith Ravi (Google), and Stefan Riezler (Heidelberg).

We would like to thank the members of the Program Committee for their timely and thoughtful reviews.

Reza Haffari, Colin Cherry, George Foster, Shahram Khadivi, and Bahar Salehi

Organizers:

Reza Haffari, Monash University
Colin Cherry, Google Research
George Foster, Google Research
Shahram Khadivi, eBay Research
Bahar Salehi, The University of Melbourne

Program Committee:

Isabelle Augenstein, University of Copenhagen
Mohit Bansal, The University of North Carolina
Daniel Beck, The University of Melbourne
Parminder Bhatia, Amazon
Colin Cherry, Google Research
Jacob Devlin, Google Research
Kevin Duh, Johns Hopkins University
Orhan Firat, Google Research
George Foster, Google Research
Reza Haffari, Monash University
Cong Vu Hoang, The University of Melbourne
Melvin Johnson, Google Research
Shahram Khadivi, eBay Research
Philipp Koehn, Johns Hopkins University
Julia Kreutzer, Heidelberg University
Gaurav Kumar, Johns Hopkins University
Patrick Littell, Carnegie Mellon University
Evgeny Matusov, eBay Research
David Mortensen, Carnegie Mellon University
Marek Rei, University of Cambridge
Sebastian Ruder, Insight Research Centre for Data Analytics
Bahar Salehi, The University of Melbourne
Nicola Ueffing, eBay Research

Invited Speakers:

Trevor Cohn, The University of Melbourne
Sujith Ravi, Google Research
Stefan Riezler, Heidelberg University

Table of Contents

Program

Thursday, July 19, 2018

9:00–9:10 ***Opening Remarks***

9:10–9:50 *Invited Talk*
Stefan Riezler

9:50–10:30 *Invited Talk*
Sujith Ravi

10:30–11:00 ***Coffee Break***

11:00–12:40 **Oral Presentations**

11:00–11:25 *Phrase-Based & Neural Unsupervised Machine Translation*
Guillaume Lample, Myle Ott, Alexis Conneau, Ludovic Denoyer and Marc'Aurelio Ranzato

11:25–11:50 *Character-level Supervision for Low-resource POS Tagging*
Katharina Kann, Johannes Bjerva, Isabelle Augenstein, Barbara Plank and Anders Søgaard

11:50–12:15 *Training a Neural Network in a Low-Resource Setting on Automatically Annotated Noisy Data*
Michael A. Hedderich and Dietrich Klakow

12:15–12:40 *Exploiting Cross-Lingual Subword Similarities in Low-Resource Document Classification*
Mozhi Zhang, Yoshinari Fujinuma and Jordan Boyd-Graber

12:40–14:00 ***Lunch Break***

14:00–15:30 Poster Session

Multi-task learning for historical text normalization: Size matters
Marcel Bollmann, Anders Søgaard and Joachim Bingel

Compositional Language Modeling for Icon-Based Augmentative and Alternative Communication
Shiran Dudy and Steven Bedrick

Multimodal Neural Machine Translation for Low-resource Language Pairs using Synthetic Data
Koel Dutta Chowdhury, Mohammed Hasanuzzaman and Qun Liu

Morphological neighbors beat word2vec on the long tail
Clayton Greenberg, Mittul Singh and Dietrich Klakow

Multi-Task Active Learning for Neural Semantic Role Labeling on Low Resource Conversational Corpus
Fariz Ikhwantri, Samuel Louvan, Kemal Kurniawan, Bagas Abisena, Valdi Rachman, Alfan Farizki Wicaksono and Rahmad Mahendra

Domain Adapted Word Embeddings for Improved Sentiment Classification
Prathusha Kameswara Sarma, Yingyu Liang and Bill Sethares

Investigating Effective Parameters for Fine-tuning of Word Embeddings Using Only a Small Corpus
Kanako Komiya and Hiroyuki Shinnou

Dependency Parsing of Code-Switching Data with Cross-Lingual Feature Representations
KyungTae Lim, Niko Partanen, Michael Rießler and Thierry Poibeau

Semi-Supervised Learning with Auxiliary Evaluation Component for Large Scale e-Commerce Text Classification
Mingkuan Liu, Musen Wen, Selcuk Kopru, Xianjing Liu and Alan Lu

Low-rank passthrough neural networks
Antonio Valerio Miceli Barone

Embedding Transfer for Low-Resource Medical Named Entity Recognition: A Case Study on Patient Mobility
Denis Newman-Griffis and Ayah Zirikly

Thursday, July 19, 2018 (continued)

15:30–16:00 ***Coffee Break***

16:00–16:40 *Invited Talk*
Trevor Cohn

16:40–17:40 ***Panel Discussion***

17:40–17:55 ***Closing Remarks***