

RepEval 2017

**The 2nd Workshop on Evaluating Vector-Space
Representations for NLP**

Proceedings of the Workshop

September 8, 2017
Copenhagen, Denmark

©2017 The Association for Computational Linguistics

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 978-1-945626-00-5

Introduction

This workshop deals with the evaluation of general-purpose vector representations for linguistic units (morphemes, words, phrases, sentences, etc). What distinguishes these representations (or embeddings) is that they are not trained with a specific application in mind, but rather to capture broadly useful features of the represented units. Another way to view their usage is through the lens of transfer learning: The embeddings are trained with one objective, but applied on others.

Evaluating general-purpose representation learning systems is fundamentally difficult. They can be trained on a variety of objectives, making simple intrinsic evaluations useless as a means of comparing methods. They are also meant to be applied to a variety of downstream tasks, which will place different demands on them, making no single extrinsic evaluation definitive. The best techniques for evaluating embedding methods in downstream tasks often require investing considerable time and resources in retraining large neural network models, making broad suites of downstream evaluations impractical. In many cases, especially for word-level embeddings, these constraints have led to the rise of dedicated evaluation tasks like similarity and analogy which are not directly related either to training objectives or to downstream tasks. Tasks like these can serve a valuable role in principle, but in practice performance on these tasks has not been highly predictive of downstream task performance.

This workshop aims foster discussion of these issues, and to support the search for high-quality general purpose representation learning techniques for NLP. The workshop will accept submissions through two tracks: a proposal track will showcase submitted proposals for new evaluation techniques, and a shared task will accept submissions of new general purpose sentence representation systems – for which standard evaluations are notably absent – which will be evaluated on a sentence understanding task.

Organizers:

Sam Bowman, New York University
Yoav Goldberg, Bar-Ilan University
Felix Hill, Google DeepMind
Angeliki Lazaridou, Google DeepMind
Omer Levy, University of Washington
Roi Reichart, Technion – Israel Institute of Technology
Anders Søgaard, University of Copenhagen

Program Committee:

Omri Abend
Mohit Bansal
Jose Camacho Collados
Billy Chiu
Allyson Ettinger
Sahar Ghannay
Anna Gladkova
Mohit Iyyer
Douwe Kiela
Arne Kohn
Andras Kornai
Farhana Liza
Oren Melamud
Dmitrijs Milajevs
Diarmuid O’Seaghdha
Marek Rei
Laura Rimell
Naomi Saphra
Roy Schwartz
Gabriel Stanovsky
Pontus Stenetorp
Karl Stratos
Yulia Tsvetkov
Ivan Vulic
Torsten Zesch

Invited Speaker:

Yejin Choi, University of Washington
Kyunghyun Cho, New York University
Jakob Uszkoreit, Google

Table of Contents

<i>Traversal-Free Word Vector Evaluation in Analogy Space</i>	
Xiaoyin Che, Nico Ring, Willi Raschkowski, Haojin Yang and Christoph Meinel	1
<i>Hypothesis Testing based Intrinsic Evaluation of Word Embeddings</i>	
Nishant Gurnani	6
<i>Evaluation of word embeddings against cognitive processes: primed reaction times in lexical decision and naming tasks</i>	
Jeremy Auguste, Arnaud Rey and Benoit Favre	11
<i>Playing with Embeddings : Evaluating embeddings for Robot Language Learning through MUD Games</i>	
Anmol Gulati and Kumar Krishna Agrawal	17
<i>Recognizing Textual Entailment in Twitter Using Word Embeddings</i>	
Octavia-Maria Şulea	21
<i>Recurrent Neural Network-Based Sentence Encoder with Gated Attention for Natural Language Inference</i>	
Qian Chen, Xiaodan Zhu, Zhen-Hua Ling, Si Wei, Hui Jiang and Diana Inkpen	26
<i>Shortcut-Stacked Sentence Encoders for Multi-Domain Inference</i>	
Yixin Nie and Mohit Bansal	31
<i>Character-level Intra Attention Network for Natural Language Inference</i>	
Han Yang, Marta R. Costa-jussà and José A. R. Fonollosa	36
<i>Refining Raw Sentence Representations for Textual Entailment Recognition via Attention</i>	
Jorge Balazs, Edison Marrese-Taylor, Pablo Loyola and Yutaka Matsuo	41
<i>LCT-MALTA's Submission to RepEval 2017 Shared Task</i>	
Hoa Vu	46

Conference Program

Friday 8 September 2017

09:00 Opening Remarks

09:20–09:55 Shared task report

The RepEval 2017 Shared Task: Multi-Genre Natural Language Inference with Sentence Representations

Nikita Nangia, Adina Williams, Angeliki Lazaridou and Samuel Bowman

09:55–10:30 Yejin Choi (University of Washington)

10:30–11:00 Coffee Break (set up posters)

11:00–11:35 Jakob Uszkoreit (Google Research)

11:35–12:10 Kyunghyun Cho (New York University)

12:10–12:30 Few Minutes Madness (Evaluation Proposals)

Traversal-Free Word Vector Evaluation in Analogy Space

Xiaoyin Che, Nico Ring, Willi Raschkowski, Haojin Yang and Christoph Meinel

Hypothesis Testing based Intrinsic Evaluation of Word Embeddings

Nishant Gurnani

Evaluation of word embeddings against cognitive processes: primed reaction times in lexical decision and naming tasks

Jeremy Auguste, Arnaud Rey and Benoit Favre

Playing with Embeddings : Evaluating embeddings for Robot Language Learning through MUD Games

Anmol Gulati and Kumar Krishna Agrawal

Friday 8 September 2017 (continued)

Recognizing Textual Entailment in Twitter Using Word Embeddings

Octavia-Maria Şulea

12:30–14:00 **Lunch (somewhere together if pos)**

14:00–14:30 **Contributed Talks (shared task systems)**

14:00–14:15 *Recurrent Neural Network-Based Sentence Encoder with Gated Attention for Natural Language Inference*

Qian Chen, Xiaodan Zhu, Zhen-Hua Ling, Si Wei, Hui Jiang and Diana Inkpen

14:15–14:30 *Shortcut-Stacked Sentence Encoders for Multi-Domain Inference*

Yixin Nie and Mohit Bansal

14:30–15:30 **Posters and discussion**

Character-level Intra Attention Network for Natural Language Inference

Han Yang, Marta R. Costa-jussà and José A. R. Fonollosa

Refining Raw Sentence Representations for Textual Entailment Recognition via Attention

Jorge Balazs, Edison Marrese-Taylor, Pablo Loyola and Yutaka Matsuo

LCT-MALTA's Submission to RepEval 2017 Shared Task

Hoa Vu

Friday 8 September 2017 (continued)

15:30–16:00 Working Coffee Break

16:00–17:30 Presentation of Findings and Panel Discussion

Yejin Choi, Kyunghyun Cho, Jakob Uszkoreit and other great minds if they are up for it...

