ACL-IJCNLP 2015

The 3rd Workshop on Continuous Vector Space Models and their Compositionality (CVSC)

Proceedings of the Workshop

July 26-31, 2015
Beijing, China

Production and Manufacturing by
Taberg Media Group AB
Box 94, 562 02 Taberg
Sweden

Microsoft Research

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 978-1-932432-66-4 / 1-932432-66-3 (Volume 1)
ISBN 978-1-932432-67-1 / 1-932432-67-1 (Volume 2)

Introduction

In most natural language processing applications the goal is to extract linguistic information from raw text or to transform linguistic observations into an alternative form, for instance from speech to text or one language to another. All of these applications often involve statistical models that rely heavily on a discrete representation of linguistic concepts, such as words and their POS tags, phrases and their syntactic categories or sentences, documents, etc. This includes any model parameterized with large probability tables or based on the extraction of multiple co-occurrence features (e.g. bigrams or tag/word pairs). Such a representation poorly models statistical structure not explicitly represented within the parameterization, but that might be very relevant from a morphological, syntactic and semantic standpoint. This hinders the generalization power of the model and reduces its ability to adapt to other domains. Another consequence is that such statistical models can only model very restricted contexts of text, without suffering from the sparsity of data. For instance, even the Google n-gram corpus only includes grams up to length 5. Data sparsity is a well-known and fundamental issue in statistical NLP, for which the existing remedies based on smoothing techniques can be insufficient.

In recent years, there has been a growing interest in algorithms that learn a continuous representation for words, phrases, or documents. For instance, one can see latent semantic analysis and latent Dirichlet allocation as a mapping of documents or words into a continuous lower dimensional topic-space. Another example, continuous word vector-space models, represent word meanings with vectors that capture semantic and syntactic information. These representations can be used to induce similarity measures by computing distances between the vectors, leading to many useful applications, such as information retrieval, search query expansions, document classification and question answering.

The idea of continuous vector spaces for language modeling has been used to develop neural language models that have reached state of the art performance in several applications. Another different trend of research on continuous vector space models belongs to the family of spectral methods developed to overcome some limitations of discrete latent space models. A further line of research is distributional semantic models that are historically more tied with linguistic theories.

Despite the success of single word vector space models, they are severely limited since they do not capture compositionality, the important quality of natural language that allows speakers to determine the meaning of a longer expression based on the meanings of its words and the rules used to combine them. This prevents them from gaining a deeper understanding of the semantics of longer phrases or sentences. For this reason, recently, there has been much progress in capturing compositionality in vector spaces.

Given this background, the first and second workshops on Continuous Vector Space Models and their Compositionality have been organized, and received high quality contributions, as well as high participation. As a result, this workshop has successfully served as a bridge between communities working on the different kinds of semantics models mentioned above. We believe, further progress on the applications of such models could be made by gathering both applied and theoretical researchers interested in developing systems that capture natural language semantics. This broader vision is reflected by the new list of topics this year. In addition, given the high interest and rapid development of various continuous semantic models, we invited more keynote speakers, compared to previous years.

This year, we continued in this direction, with the following list of topics:

- Neural networks
- Spectral methods
- Distributional semantic models

- Language modeling for automatic speech recognition, statistical machine translation, and information retrieval
- Automatic annotation of texts
- Unsupervised and semi-supervised word sense induction and disambiguation
- Phrase and sentence-level distributional representations
- The role of syntax in compositional models
- Formal and distributional semantic models
- Language modeling for logical and natural reasoning
- Integration of distributional representations with other models
- Multi-modal learning for distributional representations
- Knowledge base embedding

In brief, we aimed to highlight the ongoing effort to address some of these points, either by theoretical reasoning, or through example via demonstrating interesting properties of new or existing distributional models of semantics. We also aimed to gather formal semanticists, computational linguists, machine learning researchers and computational neuroscientists to tackle these fascinating problems.

Organizers:

Alexandre Allauzen, LIMSI-CNRS/Université Paris-Sud
Edward Grefenstette, Google DeepMind
Karl Moritz Hermann, Google DeepMind
Hugo Larochelle, Université de Sherbrooke
Scott Wen-tau Yih, Microsoft Research

Organizers:

Alexandre Allauzen, LIMSI-CNRS/Université Paris-Sud
Edward Grefenstette, Google DeepMind
Karl Moritz Hermann, Google DeepMind
Hugo Larochelle, Université de Sherbrooke
Scott Wen-tau Yih, Microsoft Research

Program Committee:

Marco Baroni, University of Trento
Yoshua Bengio, Université de Montreal
Phil Blunsom, University of Oxford
Antoine Bordes, Facebook
Leon Bottou, Microsoft
Stephen Clark, University of Cambridge
Shay Cohen, University of Edinburgh
Georgiana Dinu, University of Trento
Kevin Duh, Nara Institute of Science and Technology
Andriy Mnih, Google DeepMind
Mehrnoosh Sadrzadeh, University of London
Mark Steedman, University of Edinburgh
Peter Turney, NRC
Jason Weston, Facebook
Guillaume Wisniewski, LIMSI-CNRS

Invited Speaker:

Kyunghyun Cho, Université de Montréal
Stephen Clark, University of Cambridge
Yoav Goldberg, Bar Ilan University
Ray Mooney, University of Texas at Austin
Percy Liang, Stanford University

Table of Contents

Conference Program

Friday, July 31, 2015

09:00–09:05 *Opening Remarks*

09:05–09:50 *INVITED TALK - Kyunghyun Cho*

09:50–10:10 *CONTRIBUTED TALK - Observed versus latent features for knowledge base and text inference, Kristina Toutanova and Danqi Chen*

10:10–10:30 *CONTRIBUTED TALK - Learning Embeddings for Transitive Verb Disambiguation by Implicit Tensor Factorization, Kazuma Hashimoto and Yoshimasa Tsuruoka*

10:30–11:00 *Coffee Break*

11:00–11:45 *INVITED TALK - Yoav Goldberg*

11:45–12:30 *INVITED TALK - Percy Liang*

12:30–14:00 *Lunch*

14:00–14:45 *INVITED TALK - Stephen Clark*

14:45–15:30 *INVITED TALK - Ray Mooney*

15:30–16:00 *Coffee Break*

16:00–17:00 *Poster session*

Learning Embeddings for Transitive Verb Disambiguation by Implicit Tensor Factorization
Kazuma Hashimoto and Yoshimasa Tsuruoka

Friday, July 31, 2015 (continued)

Recursive Neural Networks Can Learn Logical Semantics
Samuel R. Bowman, Christopher Potts and Christopher D. Manning

Concept Extensions as the Basis for Vector-Space Semantics: Combining Distributional and Ontological Information about Entities
Jackie Chi Kit Cheung

Joint Semantic Relevance Learning with Text Data and Graph Knowledge
Dongxu Zhang, Bin Yuan, Dong Wang and Rong Liu

Exploring the effect of semantic similarity for Phrase-based Machine Translation
Kunal Sachdeva and Dipti Sharma

Incremental Adaptation Strategies for Neural Network Language Models
Alex Ter-Sarkisov, Holger Schwenk, Fethi Bougares and Loïc Barrault

Observed versus latent features for knowledge base and text inference
Kristina Toutanova and Danqi Chen

17:00–17:30 ***Panel***