

ACL 2013

**51st Annual Meeting of the  
Association for Computational Linguistics**

**Proceedings of the Workshop on Continuous Vector Space  
Models and their Compositionality**

August 9, 2013  
Sofia, Bulgaria

Production and Manufacturing by  
*Omnipress, Inc.*  
*2600 Anderson Street*  
*Madison, WI 53704 USA*

©2013 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](mailto:acl@aclweb.org)

ISBN: 978-1-937284-67-1

## Introduction

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 (Landauer and Dumais, 1997) and latent Dirichlet allocation (Blei et al. 2003) as a mapping of documents or words into a continuous lower dimensional topic-space. Another example, continuous word vector-space models (Sahlgren 2006, Reisinger 2012, Turian et al., 2010, Huang et al., 2012) 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 (Schuetze 1992, Manning et al., 2008), search query expansions (Jones et al., 2006), document classification (Sebastiani, 2002) and question answering (Tellex et al., 2003).

On the fundamental task of language modeling, many hard clustering approaches have been proposed such as Brown clustering (Brown et al., 1992) or exchange clustering (Martin et al., 1998). These algorithms can provide desparsification and can be seen as examples of unsupervised pre-training. However, they have not been shown to consistently outperform models based on Kneser-Ney smoothed language models which have at their core discrete n-gram representations. On the contrary, one influential proposal that uses the idea of continuous vector spaces for language modeling is that of neural language models (Bengio et al., 2003, Mikolov 2012). In these approaches, n-gram probabilities are estimated using a continuous representation of words in lieu of standard discrete representations, using a neural network that performs both the projection and the probability estimate. They report state of the art performance on several well studied language modeling datasets.

Other neural network based models that use continuous vector representations achieve state of the art performance in speech recognition applications (Schwenk, 2007, Dahl et al. 2011), multitask learning, NER and POS tagging (Collobert et al., 2011) or sentiment analysis (Socher et al. 2011). Moreover, in (Le et al., 2012), a continuous space translation model was introduced and its use in a large scale machine translation system yielded promising results in the last WMT evaluation.

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 (Frege, 1892). This prevents them from gaining a deeper understanding of the semantics of longer phrases or sentences. Recently, there has been much progress in capturing compositionality in vector spaces, e.g., (Pado and Lapata 2007; Erk and Pado 2008; Mitchell and Lapata, 2010; Baroni and Zamparelli, 2010; Zanzotto et al., 2010; Yessenalina and Cardie, 2011; Grefenstette and Sadrzadeh 2011). The work of Socher et al. 2012 compares several of these approaches on supervised tasks and for phrases of arbitrary type and length.

Another different trend of research belongs to the family of spectral methods. The motivation in that context is that working in a continuous space allows for the design of algorithms that are not plagued with the local minima issues that discrete latent space models (e.g. HMM trained with EM) tend to suffer from (Hsu et al. 2008). In fact, this motivation strikes with the conventional justification behind vector space models from the neural network literature, which are usually motivated as a way of tackling data sparsity issues. This apparent dichotomy is interesting and has not been investigated yet. Finally, spectral methods have recently been developed for word representation learning (Dhillon et al. 2011), dependency parsing (Dhillon et al. 2012) and probabilistic context-free grammars (Cohen et al. 2012).

In this workshop, we bring together researchers who are interested in how to learn continuous vector space models, their compositionality and how to use this new kind of representation in NLP applications. The goal is to review the recent progress and propositions, to discuss the challenges, to identify promising future research directions and the next challenges for the NLP community.



**Organizers:**

Alexandre Allauzen, LIMSI-CNRS/Université Paris-Sud  
Hugo Larochelle, Université de de Sherbrooke  
Richard Socher, Stanford University  
Christopher Manning, Stanford University

**Program Committee:**

Yoshua Bengio, Université de Montréal  
Antoine Bordes, Université Technologique de Compiègne  
Léon Bottou, Microsoft Research  
Xavier Carreras, Universitat Politècnica de Catalunya  
Shay Cohen, Columbia University  
Michael Collins, Columbia University  
Ronan Collobert, IDIAP Research Institute  
Kevin Duh, Nara Institute of Science and Technology  
Dean Foster, University of Pennsylvania  
Percy Liang, Stanford University  
Andriy Mnih, Gatsby Computational Neuroscience Unit  
John Platt, Microsoft Research  
Holger Schwenk, Université du Maine  
Jason Weston, Google  
Guillaume Wisniewski, LIMSI-CNRS/Université

**Invited Speaker:**

Xavier Carreras, Universitat Politècnica de Catalunya  
Mirella Lapata, University of Edinburgh



## Table of Contents

<i>Vector Space Semantic Parsing: A Framework for Compositional Vector Space Models</i> Jayant Krishnamurthy and Tom Mitchell .....	1
<i>Learning from errors: Using vector-based compositional semantics for parse reranking</i> Phong Le, Willem Zuidema and Remko Scha .....	11
<i>A Structured Distributional Semantic Model : Integrating Structure with Semantics</i> Kartik Goyal, Sujay Kumar Jauhar, Huiying Li, Mrinmaya Sachan, Shashank Srivastava and Eduard Hovy .....	20
<i>Letter N-Gram-based Input Encoding for Continuous Space Language Models</i> Henning Sperr, Jan Niehues and Alex Waibel .....	30
<i>Transducing Sentences to Syntactic Feature Vectors: an Alternative Way to "Parse"?</i> Fabio Massimo Zanzotto and Lorenzo Dell' Arciprete .....	40
<i>General estimation and evaluation of compositional distributional semantic models</i> Georgiana Dinu, Nghia The Pham and Marco Baroni .....	50
<i>Applicative structure in vector space models</i> Marton Makrai, David Mark Nemeskey and Andras Kornai .....	59
<i>Determining Compositionality of Expressions Using Various Word Space Models and Methods</i> Lubomír Krčmář, Karel Ježek and Pavel Pecina .....	64
<i>"Not not bad" is not "bad": A distributional account of negation</i> Karl Moritz Hermann, Edward Grefenstette and Phil Blunsom .....	74
<i>Towards Dynamic Word Sense Discrimination with Random Indexing</i> Hans Moen, Erwin Marsi and Björn Gambäck .....	83
<i>A Generative Model of Vector Space Semantics</i> Jacob Andreas and Zoubin Ghahramani .....	91
<i>Aggregating Continuous Word Embeddings for Information Retrieval</i> Stephane Clinchant and Florent Perronnin .....	100
<i>Answer Extraction by Recursive Parse Tree Descent</i> Christopher Malon and Bing Bai .....	110
<i>Recurrent Convolutional Neural Networks for Discourse Compositionality</i> Nal Kalchbrenner and Phil Blunsom .....	119





# Conference Program

## (9:00) Oral session 1

- 9:00 Opening
- 9:05 INVITED TALK: Structured Prediction with Low-Rank Bilinear Models by Xavier Carreras
- 10:00 *Vector Space Semantic Parsing: A Framework for Compositional Vector Space Models*  
Jayant Krishnamurthy and Tom Mitchell
- 10:20 *Learning from errors: Using vector-based compositional semantics for parse reranking*  
Phong Le, Willem Zuidema and Remko Scha
- 10:40 Coffee Break

## (11:00) Poster session

*A Structured Distributional Semantic Model : Integrating Structure with Semantics*  
Kartik Goyal, Sujay Kumar Jauhar, Huiying Li, Mrinmaya Sachan, Shashank Srivastava and Eduard Hovy

*Letter N-Gram-based Input Encoding for Continuous Space Language Models*  
Henning Sperr, Jan Niehues and Alex Waibel

*Transducing Sentences to Syntactic Feature Vectors: an Alternative Way to "Parse"?*  
Fabio Massimo Zanzotto and Lorenzo Dell'Arciprete

*General estimation and evaluation of compositional distributional semantic models*  
Georgiana Dinu, Nghia The Pham and Marco Baroni

*Applicative structure in vector space models*  
Marton Makrai, David Mark Nemeskey and Andras Kornai

*Determining Compositionality of Expressions Using Various Word Space Models and Methods*  
Lubomír Krčmář, Karel Ježek and Pavel Pecina

*"Not not bad" is not "bad": A distributional account of negation*  
Karl Moritz Hermann, Edward Grefenstette and Phil Blunsom

**Poster session (continued)**

*Towards Dynamic Word Sense Discrimination with Random Indexing*

Hans Moen, Erwin Marsi and Björn Gambäck

12:30 Lunch Break

**(14:00) Oral session 2**

14:00 INVITED TALK: Learning to Ground Meaning in the Visual World by Mirella Lapata

15:00 *A Generative Model of Vector Space Semantics*

Jacob Andreas and Zoubin Ghahramani

15:20 *Aggregating Continuous Word Embeddings for Information Retrieval*

Stephane Clinchant and Florent Perronnin

15:40 Coffee Break

16:00 *Answer Extraction by Recursive Parse Tree Descent*

Christopher Malon and Bing Bai

16:20 *Recurrent Convolutional Neural Networks for Discourse Compositionality*

Nal Kalchbrenner and Phil Blunsom

**(16:40) Panel Discussion**