

NAACL HLT 2015

**The 1st Workshop on Vector Space
Modeling for Natural Language
Processing (in NAACL 2015)**

Proceedings of the Workshop

June 5, 2015
Denver, Colorado, USA

©2015 The Association for Computational Linguistics

Order print-on-demand copies from:

Curran Associates
57 Morehouse Lane
Red Hook, New York 12571
USA
Tel: +1-845-758-0400
Fax: +1-845-758-2633
curran@proceedings.com

ISBN 978-1-941643-46-4

Introduction

The idea of statistical analysis of language is an old idea, but modern NLP started with a focus on methods based on pure symbolic analysis of language. Statistical methods were introduced to NLP in its current form in the 1980s/1990s, allowing “soft” reasoning about language, and made NLP more data-driven. Over the last decade another step has been taken in this direction – it was proposed to represent and analyze language in vector spaces. Now-a-days, context, symbolic and high-dimensional representations are often augmented with relatively low-dimensional vector-space representations. Vector space representations have been successfully used in different areas of NLP such as syntax and semantics.

This workshop is an opportunity to explore state of the art in the use of vector spaces in order to computationally analyze natural language. The focus of the workshop is on the use of vector spaces to learn latent representations.

The goal of the workshop is to bring together researchers from areas such as deep learning and representation learning, spectral learning, distributional compositional semantics and others, in order to see their relevance to each other, and learn about the state of the art in these areas.

This is the first time that this workshop is held. There were other similar workshops in the past, such as the Workshop on Continuous Vector Space Models and their Compositionality.

The program this year includes 27 papers that cover different areas under the realm of vector space modeling in NLP, all of which are presented in two poster sessions. There are also 3 invited speakers, Marco Baroni, Chris Manning and Xavier Carreras, with each of their talks covering a different aspect of vector space modeling in NLP.

We would like to thank the Program Committee members who reviewed the papers this year. We would also like to thank the workshop participants. Last, a word of thanks also goes to our two sponsors: Google Deepmind and Textkernel.

Phil Blunsom, Shay Cohen, Paramveer Dhillon and Percy Liang

Co-Organizers

Organizers:

Phil Blunsom (University of Oxford)
Shay Cohen (University of Edinburgh)
Paramveer Dhillon (Massachusetts Institute of Technology)
Percy Liang (Stanford University)

Program Committee:

Waleed Ammar (Carnegie Mellon University)
Apoorv Agarwal (Columbia University)
Borja Balle (McGill University)
William Blacoe (University of Edinburgh)
Jordan Boyd-Graber (University of Colorado)
Xavier Carreras (XRCE)
Tejaswini Deoskar (University of Edinburgh)
Weisi Duan (Carnegie Mellon University)
Edward Grefenstette (University of Oxford)
Manaal Frauqui (Carnegie Mellon University)
Lea Frermann (University of Edinburgh)
Karl Moritz Hermann (University of Oxford)
Thang Minh Luong (Stanford University)
Annie Louis (University of Edinburgh)
Jeff Mitchell (University of Edinburgh)
Ankur Parikh (Carnegie Mellon University)
Rohan Ramanath (Carnegie Mellon University)
Siva Reddy (University of Edinburgh)
Roi Reichart (Technion)
Sebastian Riedel (University College London)
Richard Socher (Stanford University)
Karl Stratos (Columbia University)
Lyle Ungar (University of Pennsylvania)
Eva Maria Vecchi (University of Cambridge)

Invited Speakers:

Marco Baroni (University of Trento)
Xavier Carreras (XRCE)
Chris Manning (Stanford University)

Sponsors:

Google DeepMind

Textkernel (<http://www.textkernel.com>: machine learning for matching people and jobs)

Table of Contents

<i>A Simple Word Embedding Model for Lexical Substitution</i>	
Oren Melamud, Omer Levy and Ido Dagan	1
<i>Unsupervised Text Normalization Using Distributed Representations of Words and Phrases</i>	
Vivek Kumar Rangarajan Sridhar	8
<i>A Multi-classifier Approach to support Coreference Resolution in a Vector Space Model</i>	
Ana Zelaia, Olatz Arregi and Basilio Sierra	17
<i>Neural context embeddings for automatic discovery of word senses</i>	
Mikael Kågebäck, Fredrik Johansson, Richard Johansson and Devdatt Dubhashi	25
<i>Distributional Representations of Words for Short Text Classification</i>	
Chenglong Ma, Weiqun Xu, Peijia Li and Yonghong Yan	33
<i>Relation Extraction: Perspective from Convolutional Neural Networks</i>	
Thien Huu Nguyen and Ralph Grishman	39
<i>Distributional Semantic Concept Models for Entity Relation Discovery</i>	
Jay Urbain, Glenn Bushee and George Kowalski	49
<i>A Deep Architecture for Non-Projective Dependency Parsing</i>	
Erick Fonseca and Sandra Aluísio	56
<i>Short Text Clustering via Convolutional Neural Networks</i>	
jiaming xu, peng wang, guanhua tian, bo xu, jun zhao, fangyuan wang and hongwei hao	62
<i>A Word-Embedding-based Sense Index for Regular Polysemy Representation</i>	
Marco Del Tredici and Núria Bel	70
<i>Simple Semi-Supervised POS Tagging</i>	
Karl Stratos and Michael Collins	79
<i>Learning Distributed Representations for Multilingual Text Sequences</i>	
Hieu Pham, Thang Luong and Christopher Manning	88
<i>Combining Distributed Vector Representations for Words</i>	
Justin Garten, Kenji Sagae, Volkan Ustun and Morteza Dehghani	95
<i>Dependency Link Embeddings: Continuous Representations of Syntactic Substructures</i>	
Mohit Bansal	102
<i>DeepNL: a Deep Learning NLP pipeline</i>	
Giuseppe Attardi	109
<i>A Vector Space Approach for Aspect Based Sentiment Analysis</i>	
Abdulaziz Alghunaim, Mitra Mohtarami, Scott Cyphers and Jim Glass	116

<i>Word Embeddings vs Word Types for Sequence Labeling: the Curious Case of CV Parsing</i> Melanie Tosik, Carsten Lygteskov Hansen, Gerard Goossen and Mihai Rotaru	123
<i>Morpho-syntactic Regularities in Continuous Word Representations: A multilingual study.</i> Garrett Nicolai, Colin Cherry and Grzegorz Kondrak	129
<i>Towards Combined Matrix and Tensor Factorization for Universal Schema Relation Extraction</i> Sameer Singh, Tim Rocktäschel and Sebastian Riedel	135
<i>Neural word embeddings with multiplicative feature interactions for tensor-based compositions</i> Joo-Kyung Kim, Marie-Catherine de Marneffe and Eric Fosler-Lussier	143
<i>Bilingual Word Representations with Monolingual Quality in Mind</i> Thang Luong, Hieu Pham and Christopher D. Manning	151
<i>Distributed Word Representations Improve NER for e-Commerce</i> Mahesh Joshi, Ethan Hart, Mirko Vogel and Jean-David Ruvini	160
<i>Semantic Information Extraction for Improved Word Embeddings</i> Jiaqiang Chen and Gerard de Melo	168
<i>Named Entity Recognition for Arabic Social Media</i> Ayah Zirikly and Mona Diab	176
<i>Vector Space Models for Scientific Document Summarization</i> John Conroy and Sashka Davis	186
<i>Unsupervised Topic Modeling for Short Texts Using Distributed Representations of Words</i> Vivek Kumar Rangarajan Sridhar	192
<i>Estimating User Location in Social Media with Stacked Denoising Auto-encoders</i> Ji Liu and Diana Inkpen	201

Conference Program

9:00–9:15 *Opening Remarks*

9:15–10:15 *Invited Talk (Chris Manning)*

10:15–12:15 *Poster session 1*

A Simple Word Embedding Model for Lexical Substitution

Oren Melamud, Omer Levy and Ido Dagan

Unsupervised Text Normalization Using Distributed Representations of Words and Phrases

Vivek Kumar Rangarajan Sridhar

A Multi-classifier Approach to support Coreference Resolution in a Vector Space Model

Ana Zelaia, Olatz Arregi and Basilio Sierra

Neural context embeddings for automatic discovery of word senses

Mikael Kågebäck, Fredrik Johansson, Richard Johansson and Devdatt Dubhashi

Distributional Representations of Words for Short Text Classification

Chenglong Ma, Weiqun Xu, Peijia Li and Yonghong Yan

Relation Extraction: Perspective from Convolutional Neural Networks

Thien Huu Nguyen and Ralph Grishman

Distributional Semantic Concept Models for Entity Relation Discovery

Jay Urbain, Glenn Bushee and George Kowalski

A Deep Architecture for Non-Projective Dependency Parsing

Erick Fonseca and Sandra Aluísio

Short Text Clustering via Convolutional Neural Networks

jiaming xu, peng wang, guanhua tian, bo xu, jun zhao, fangyuan wang and hongwei hao

A Word-Embedding-based Sense Index for Regular Polysemy Representation

Marco Del Tredici and Núria Bel

No Day Set (continued)

Simple Semi-Supervised POS Tagging

Karl Stratos and Michael Collins

Learning Distributed Representations for Multilingual Text Sequences

Hieu Pham, Thang Luong and Christopher Manning

Combining Distributed Vector Representations for Words

Justin Garten, Kenji Sagae, Volkan Ustun and Morteza Dehghani

12:15–13:30 *Lunch*

13:30–14:30 *Invited Talk (Xavier Carreras)*

14:30–16:30 *Poster session 2*

Dependency Link Embeddings: Continuous Representations of Syntactic Substructures

Mohit Bansal

DeepNL: a Deep Learning NLP pipeline

Giuseppe Attardi

A Vector Space Approach for Aspect Based Sentiment Analysis

Abdulaziz Alghunaim, Mitra Mohtarami, Scott Cyphers and Jim Glass

Word Embeddings vs Word Types for Sequence Labeling: the Curious Case of CV Parsing

Melanie Tosik, Carsten Lygteskov Hansen, Gerard Goossen and Mihai Rotaru

Morpho-syntactic Regularities in Continuous Word Representations: A multilingual study.

Garrett Nicolai, Colin Cherry and Grzegorz Kondrak

Towards Combined Matrix and Tensor Factorization for Universal Schema Relation Extraction

Sameer Singh, Tim Rocktäschel and Sebastian Riedel

Neural word embeddings with multiplicative feature interactions for tensor-based compositions

Joo-Kyung Kim, Marie-Catherine de Marneffe and Eric Fosler-Lussier

No Day Set (continued)

Bilingual Word Representations with Monolingual Quality in Mind

Thang Luong, Hieu Pham and Christopher D. Manning

Distributed Word Representations Improve NER for e-Commerce

Mahesh Joshi, Ethan Hart, Mirko Vogel and Jean-David Ruvini

Semantic Information Extraction for Improved Word Embeddings

Jiaqiang Chen and Gerard de Melo

Named Entity Recognition for Arabic Social Media

Ayah Zirikly and Mona Diab

Vector Space Models for Scientific Document Summarization

John Conroy and Sashka Davis

Unsupervised Topic Modeling for Short Texts Using Distributed Representations of Words

Vivek Kumar Rangarajan Sridhar

Estimating User Location in Social Media with Stacked Denoising Auto-encoders

Ji Liu and Diana Inkpen

16:30–17:30 *Invited Talk (Marco Baroni)*

17:30–17:45 *Concluding Remarks and Prizes*

17:45–19:00 *Farewell Reception*

