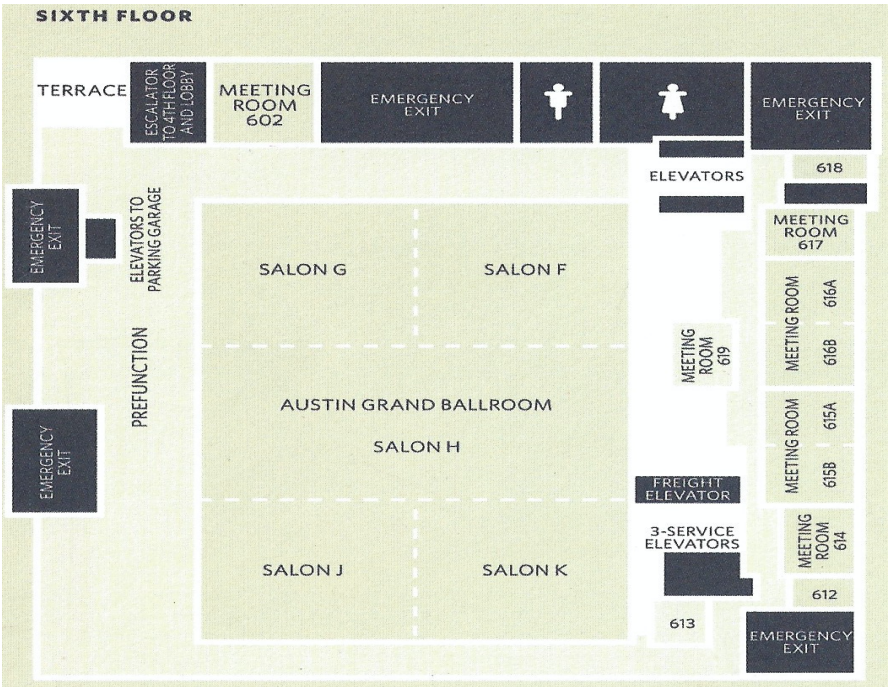
emnlp2016

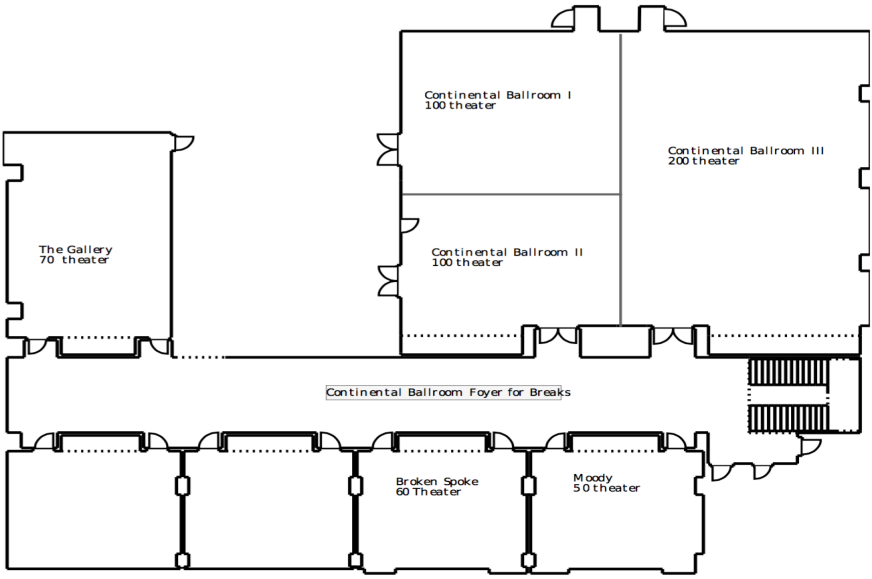
November 1–5, 2016 Austin, Texas, USA

Conference on Empirical Methods in Natural Language Processing

CONFERENCE HANDBOOK

www.emnlp2016.net


SIXTH FLOOR
TERRACE
ESCALATOR TO 4TH FLOOR AND LOBBY
MEETING ROOM 602
EMERGENCY EXIT
EMERGENCY EXIT
ELEVATORS
618
MEETING ROOM 617
EMERGENCY EXIT
ELEVATORS TO PARKING GARAGE
SALON G
SALON F
MEETING ROOM 619
MEETING ROOM 616A
MEETING ROOM 616B
PREFUNCTION
AUSTIN GRAND BALLROOM
SALON H
EMERGENCY EXIT
MEETING ROOM 615A
MEETING ROOM 615B
FREIGHT ELEVATOR
3-SERVICE ELEVATORS
SALON J
SALON K
MEETING ROOM 614
612
613
EMERGENCY EXIT


Continental Ballroom I
100 theater
Continental Ballroom III
200 theater
The Gallery
70 theater
Continental Ballroom II
100 theater
Continental Ballroom Foyer for Breaks
Broken Spoke
60 Theater
Moody
50 theater

Cover design by Aurelia Bunescu
Thanks to Matt Post & Fernando Batista for their notes on creating a handbook
Handbook assembled by Swapna Somasundaran

Meal Info

The main conference registration includes

- Welcome Reception
- Two Poster Session dinners
- Two breaks each day with snacks

WIFI access

Network: Hilton Meetings
Password: EMNLP16

Links

Web site: http://www.emnlp2016.net/
Contact email: contact@emnlp2016.net
Find us on facebook: https://www.facebook.com/EMNLP-2016-715509195252840/
Follow us on twitter: https://twitter.com/emnlp2016
Conference4me app: XXXXXXXXXXXXXXX

Contents

1 Conference Information

Preface by the General Chair

October 15, 2016

Welcome to the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP 2016) in Austin, Texas, USA!

EMNLP is annually organized by SIGDAT, the Association for Computational Linguistics' special interest group on linguistic data and corpus-based approaches to NLP. At EMNLP 2016, one of the top tier conferences in Natural Language Processing (NLP), we have witnessed how our field thrives. This is not only reflected on the number of paper submissions but also the number of sponsors. The number of long paper submissions has increased 14.5% over that of 2015. This year, we also have seen a record high number of sponsors in EMNLP history. We're honored and grateful to have Amazon, Baidu, Google and Grammarly as the Platinum Sponsors, Bloomberg, Citadel, eBay, Facebook, IBM Research, Maluuba and Microsoft as the Gold Sponsors, AI@ISI as the Silver Sponsor, Nuance, VoiceBox and Yandex as the Bronze Sponsors. We also have Oracle as the Student Volunteer Sponsor. Further more, we're expecting a record attendance.

A large number of people worked hard to bring this annual meeting to fruition, to whom I'm very grateful. Program Chairs, **Kevin Duh** and **Xavier Carreras**, the Area Chairs, reviewers, best paper committee members put immense amount of work to develop the technical program. Tutorial Chairs, **Bishan Yang** and **Rebecca Hwa**, Workshop Chairs, **Annie Louis** and **Greg Kondrak** conducted a competitive selection process in collaboration with NAACL and ACL to select 6 tutorials and 8 workshops. Sponsorship Chairs, **Michel Galley**, **Hang Li** (ACL International Sponsorship Committee Representative for EMNLP) did an excellent job to attract the record number of sponsors. Publication Chairs, **Siddharth Patwardhan**,**Daniele Pighin**(advisor), Handbook Chair, **Swapna Somasundaran** worked in very tight schedule to assemble the proceedings, C4Me Mobile app and handbooks. Publicity Chair, **Saif M. Mohammad** disseminated the call for papers, call for participation and other announcements in a timely manner. Webmaster, **Jackie C.K. Cheung** kept the website updated all the time, providing a professional outlook of the conference. Student Scholarship Chair and Student Volunteer Coordinator, **Vincent Ng**, played two critical roles, managing the NSF and SIGDAT scholarship, and the review of applications, coordinating the student volunteers to support the conference. SIGDAT Secretary, **Chris Callison-Burch** acted as the liaison between SIGDAT and the conference organizers. He is always available to provide good suggestions.

As usual, the conference cannot be done without Local Arrangements Chair, **Priscilla Rasmussen**, who single-handedly took care of all conference logistics. I would like to

mention that I benefited from last year's General Chair, **Lluís Màrquez**, for the monthly progress reports and other sharing. We are also grateful to the invited speakers, **Christopher Potts**, **Andreas Stolcke** and **Stefanie Tellex** who will share with us their exciting research.

I really appreciate the trust from SIGDAT officers, including previous secretary, **Noah Smith** for me to coordinate the conference as the General Chair.

Finally, I'd like to thank all the authors and attendees. Your participation made a difference to the conference. I hope that you have an enjoyable and productive time at Austin. My best wishes for a success conference!

Jian Su
EMNLP 2016 General Chair

Preface by the Program Committee Co-Chairs

October 17, 2016

Welcome to the 2016 Conference on Empirical Methods in Natural Language Processing! This year we received 1,087 valid submissions, of which 687 were long papers and 400 were short papers. We accepted 177 long papers (25.8% acceptance rate) and 87 short papers (21.8% acceptance rate), for a total of 264 papers and an overall acceptance rate of 24.3%.

The technical program at EMNLP 2016 consists of a total of 273 papers, including 9 journal papers accepted by the Transactions of ACL. We have structured the conference into three parallel oral sessions in the day and two poster sessions in the evening. Borrowing from recent NAACL conference innovations, we also run poster spotlight sessions (also called *HMM: Half-Minute Madness*[1]), where poster presenters of long papers have 30 seconds and one slide to advertise their work. Poster sessions are becoming larger due to the rapid growth in our field, and we believe it is important to ensure that all papers receive the exposure they deserve.

We are excited and grateful to have three distinguished speakers for our invited keynote talks. Christopher Potts (Stanford University) will present recent advances in rational speech acts and pragmatics. Andreas Stolcke (Microsoft Research) will talk about the challenges and opportunities in human-human-machine dialog. Stefanie Tellex (Brown University) will discuss novel methods and frameworks for enabling human-robot collaboration. We think that these are exciting research areas that can potentially impact—and be impacted by—the EMNLP community in the near future. We look forward to their keynotes and the conversations afterwards.

The program committee includes 823 primary reviewers and 99 secondary reviewers. The committee was structured into 12 thematic areas, handled by 41 area chairs. We are grateful to all program committee members for their effort and dedication during our tight reviewing schedule; without them we cannot create a strong high-quality program. We are also thankful for all authors who submitted papers, which overall cover a diverse range of topics.

Best paper awards were organized around three categories: best paper, best short paper, and best resource paper. The latter category was introduced at EMNLP 2015. Since resources have become central for scientific progress in our field, we would like this category of award to become a standard. The selection process was bottom-up: reviewers and area chairs suggested candidates, which were short-listed by us program chairs. Then, for each category we created a committee of experts to discuss the papers in depth, and we chaired the committees.

For **best paper**, the committee members were Stephen Clark, Hal Daumé III, Chris Dyer, and Julia Hockenmaier. The committee selected two best long papers:

- Best Paper: *Improving Information Extraction by Acquiring External Evidence with Reinforcement Learning*, by Karthik Narasimhan, Adam Yala and Regina Barzilay.

[1] Neologism coined by Joel Tetreault, our HMM chair.

	Long	Short	Total
Initial submissions	747	438	1,185
Withdrawn or rejected without review	60	38	98
Submissions reviewed	687	400	1,087
Submissions accepted	177	87	264
Acceptance rate	25.76%	21.75%	24.29%
TACL papers	9	0	9
Papers at EMNLP 2016	186	87	273
Oral talks	87	22	109
Poster presentations	99	65	164

Table 1.1: Submission statistics of EMNLP 2016

- Best Paper: *Global Neural CCG Parsing with Optimality Guarantees*, by Kenton Lee, Mike Lewis and Luke Zettlemoyer.

In addition, two papers were given an honorable mention for best paper:

- Honorable Mention for Best Paper: *Span-Based Constituency Parsing with a Structure-Label System and Provably Optimal Dynamic Oracles*, by James Cross and Liang Huang.

- Honorable Mention for Best Paper: *Sequence-to-Sequence Learning as Beam-Search Optimization*, by Sam Wiseman and Alexander M. Rush.

For **best short paper**, the committee had Stefan Riezler, Anoop Sarkar, and Noah Smith, and the award went to:

- Best Short Paper: *Learning a Lexicon and Translation Model from Phoneme Lattices*, by Oliver Adams, Graham Neubig, Trevor Cohn, Steven Bird, Quoc Truong Do and Satoshi Nakamura.

For **best resource paper**, the committee consisted of Eneko Agirre, Mirella Lapata, and Sebastian Riedel, and the award went to:

- Best Resource Paper: *SQuAD: 100,000+ Questions for Machine Comprehension of Text*, by Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev and Percy Liang.

We are grateful to the many people who helped us at various stages of the program preparation. In particular, we would like to thank:

- Jian Su and Chris Callison-Burch, who gave us advice and support throughout the whole process, not only in their capacity as program chairs of EMNLP 2015, but also as general chair of EMNLP 2016 (Jian) and SIGDAT secretary-treasurer (Chris).

- The 41 area chairs, whose expertise and dedication we relied on heavily. They selected reviewers, coordinated the review process, led discussions, and made recommendations. We owe you a favor: Yoav Artzi, Tim Baldwin, Guillaume Bouchard, Nate Chambers, Kyunghyun Cho, Michael Collins, John DeNero, Georgiana Dinu, Sanja Fidler, Alex Fraser, Kuzman Ganchev, Ed Grefenstette, Julia Hockenmaier, Dirk Hovy, Liang Huang, Ruihong Huang, Min-Yen Kan, Daisuke Kawahara, Yang Liu, Bing Liu,

André F.T. Martins, Saif Mohammad, Ray Mooney, Smaranda Muresan, Preslav Nakov, Vivi Nastase, Ariadna Quattoni, Laura Rimell, Eric Ringger, Alan Ritter, Brian Roark, David Smith, Manfred Stede, Suzanne Stevenson, Michael Strube, Joel Tetreault, Lucy Vanderwende, Dekai Wu, Wei Xu, Scott Wen-Tau Yih, and Geoff Zweig.

- Priscilla Rasmussen, our local organizer who performed amazing feats to make everything work.
- Siddharth Patwardhan and Daniele Pighin, the publication chairs.
- Joel Tetreault, Brendan O'Connor, and Courtney Napoles for organizing and chairing *the HMM sessions*.
- The session chairs: Regina Barzilay, Alexandra Birch, Phil Blunsom, Yejin Choi, Ido Dagan, Marie-Catherine de Marneffe, Katrin Erk, Pascale Fung, Alona Fyshe, Rebecca Hwa, Heng Ji, Diane Litman, Yang Liu, Lluís Màrquez, André F.T. Martins, Kathy McKeown, Raymond Mooney, Preslav Nakov, Hinrich Schütze, Thamar Solorio, Hiroya Takamura, Kristina Toutanova, Bonnie Webber, and Wei Xu.
- Jackie C.K. Cheung, who maintained the EMNLP 2016 website with up-to-date information.
- Yejin Choi, who kept us connected with the ACL Exec.
- Kristina Toutanova and Lillian Lee, who helped us regarding TACL papers.
- Janyce Wiebe, Michael Strube, and Anoop Sarkar, who provided detailed advice about chairing a program committee of a large conference at the initial planning stages of the process.
- Ani Nenkova, Owen Rambow, Katrin Erk and Noah Smith (program co-chairs of NAACL and ACL this year), with which we coordinated several aspects of the major conferences this year.
- The Softconf support team, Rich Gerber and Paolo Gai, who assisted us in using the Start Conference Manager.

On behalf of all attendees at the conference, we would also like to acknowledge the generosity of our sponsors: Amazon, Baidu, Google, Grammarly, Bloomberg, Citadel, eBay, Facebook, IBM Research, Maluuba, Microsoft, AI@ISI, Nuance, VoiceBox, Yandex, and Oracle.

Chairing the program committee of EMNLP has been a great honor and a rich scientific experience. We are grateful to SIGDAT for giving us this opportunity. And we hope that you will find the program as exciting and enjoyable as we do!

Xavier Carreras and Kevin Duh
EMNLP 2016 Program Committee Co-Chairs

Organizing Committee

General Chair

Jian Su, Institute for Infocomm Research - I2R

Program co-Chairs

Xavier Carreras, Xerox Research Centre Europe
Kevin Duh, Johns Hopkins University

Workshops co-Chairs

Annie Louis, University of Essex
Greg Kondrak, University of Alberta

Tutorial co-Chairs

Bishan Yang, Carnegie Mellon University
Rebecca Hwa, University of Pittsburgh

Publication co-Chairs

Siddharth Patwardhan, Apple
Daniele Pighin, Google Inc.

Publicity Chair

Saif M. Mohammad, National Research Council Canada

Sponsorship Chairs

Michel Galley, Microsoft Research
Hang Li, Huawei Technologies (ISC Representative for EMNLP)

Student Scholarship Chair and Student Volunteer Coordinator

Vincent Ng, University of Texas at Dallas

Handbook Chair

Swapna Somasundaran, Educational Testing Service

Website Chair

Jackie C.K. Cheung, McGill University

SIGDAT Liaison

Chris Callison-Burch, University of Pennsylvania

Local Arrangements Chair

Priscilla Rasmussen, ACL Business Manager

Program Committee

Program co-Chairs

Xavier Carreras, Xerox Research Centre Europe
Kevin Duh, Johns Hopkins University

Area Chairs

Information Extraction, Information Retrieval, and Question Answering

Nate Chambers, United States Naval Academy
Ruihong Huang, Texas A&M University
Min-Yen Kan, National University of Singapore
Alan Ritter, The Ohio State University
Scott Wen-Tau Yih, Microsoft Research

Language and Vision

Sanja Fidler, University of Toronto
Julia Hockenmaier, University of Illinois at Urbana-Champaign

Linguistic Theories and Psycholinguistics

Suzanne Stevenson, University of Toronto

Machine Learning

Guillaume Bouchard, University College London
Kyunghyun Cho, New York University
Kuzman Ganchev, Google
Ariadna Quattoni, Xerox Research Centre Europe
Eric Ringger, Facebook

Machine Translation and Multilinguality

John DeNero, UC Berkeley
Alex Fraser, Ludwig-Maximilians-Universitat Munchen
Yang Liu, Tsinghua University
Dekai Wu, Hong Kong University of Science and Technology

Segmentation, Tagging, and Parsing

Michael Collins, Columbia University
Liang Huang, Oregon State University
Daisuke Kawahara, Kyoto University
Andre Martins, Priberam Labs

Semantics

Yoav Artzi, Cornell Tech
Georgiana Dinu, IBM
Ed Grefenstette, Google DeepMind
Ray Mooney, University of Texas at Austin
Laura Rimell, University of Cambridge

Sentiment Analysis and Opinion Mining

Dirk Hovy, University of Copenhagen
Bing Liu, University of Illinois at Chicago
Saif Mohammad, National Research Council Canada

Social Media and Computational Social Science

Tim Baldwin, University of Melbourne
Smaranda Muresan, Columbia University

Spoken Lanugage Processing

Brian Roark, Google
Geoff Zweig, Microsoft Research

Summarization, Generation, Discourse, Dialogue

Manfred Stede, Potsdam University
Michael Strube, Heidelberg Institute for Theoretical Studies
Lucy Vanderwende, Microsoft Research
Wei Xu, University of Pennsylvania

Text Mining and NLP Applications

Preslav Nakov, Qatar Computing Research Institute
Vivi Nastase, FBK
David Smith, Northeastern University
Joel Tetreault, Yahoo! Labs

2

Tutorials:Tuesday, November 1

9:00 – 12:30 **Morning Tutorials**

Practical Neural Networks for NLP: From Theory to Code
Chris Dyer, Yoav Goldberg, and Graham Neubig — *Salon K*

Advanced Markov Logic Techniques for Scalable Joint Inference in NLP
Deepak Venugopal, Vibhav Gogate, and Vincent Ng — *Room 616B*

Lifelong Machine Learning for Natural Language Processing
Zhiyuan Chen and Bing Liu — *Room 616A*

12:30 – 14:00 **Lunch break**

14:00 – 17:30 **Afternoon Tutorials**

Neural Networks for Sentiment Analysis
Yue Zhang and Duy Tin Vo — *Room 616A*

Continuous Vector Spaces for Cross-language NLP Applications
Rafael E. Banchs — *Room 616B*

Methods and Theories for Large-scale Structured Prediction
Xu Sun and Yansong Feng — *Salon K*

Practical Neural Networks for NLP: From Theory to Code

Chris Dyer, Yoav Goldberg, and Graham Neubig

Tuesday, November 1 – 09:00–12:30

Salon K

This tutorial aims to bring NLP researchers up to speed with the current techniques in deep learning and neural networks, and show them how they can turn their ideas into practical implementations. We will start with simple classification models (logistic regression and multilayer perceptrons) and cover more advanced patterns that come up in NLP such as recurrent networks for sequence tagging and prediction problems, structured networks (e.g., compositional architectures based on syntax trees), structured output spaces (sequences and trees), attention for sequence-to-sequence transduction, and feature induction for complex algorithm states. A particular emphasis will be on learning to represent complex objects as recursive compositions of simpler objects. This representation will reflect characterize standard objects in NLP, such as the composition of characters and morphemes into words, and words into sentences and documents. In addition, new opportunities such as learning to embed “algorithm states" such as those used in transition-based parsing and other sequential structured prediction models (for which effective features may be difficult to engineer by hand) will be covered.

Everything in the tutorial will be grounded in code —- we will show how to program seemingly complex neural-net models using toolkits based on the computation-graph formalism. Computation graphs decompose complex computations into a DAG, with nodes representing inputs, target outputs, parameters, or (sub)differentiable functions (e.g., “tanh", “matrix multiply", and “softmax"), and edges represent data dependencies. These graphs can be run “forward” to make predictions and compute errors (e.g., log loss, squared error) and then

---

Chris Dyer (CMU) Chris’s research looks at the intersection between machine learning and natural language processing, in particular focusing on representation learning, structured models, and multilinguality. He has received best paper nominations and awards at NAACL, EMNLP, and ACL. He is the author of the CNN toolkit, which was used by the Jelinek 2016 Summer Workshop on neural machine translation.

Yoav Goldberg (Bar Ilan) Yoav’s research focuses around structure prediction problems in natural language, in particular prediction of syntactic representations, greedy methods, semi-supervised learning and robust cross-domain performance. Recently, Yoav became involved in the application of neural-network based models to NLP problems, with two core focuses: better understanding the neural network building blocks; and the use of neural networks for structured problems. Yoav wrote a primer on neural-networks method in NLP, and is responsible for the python bindings of the CNN library.

Graham Neubig (NAIST) Graham’s research focuses on machine learning methods for language and speech processing. His work puts a particular focus on machine translation and speech translation, but also covers other NLP tasks such as tagging and syntactic/semantic parsing, and methods for learning from unlabeled data such as unsupervised, semi-supervised, or active learning. He is interested in neural network models that incorporate our linguistic intuitions, and is one of the core developers of the CNN toolkit.

“backward" to compute derivatives with respect to model parameters. In particular we’ll cover the Python bindings of the CNN library. CNN has been designed from the ground up for NLP applications, dynamically structured NNs, rapid prototyping, and a transparent data and execution model.

Advanced Markov Logic Techniques for Scalable Joint Inference in NLP

Deepak Venugopal, Vibhav Gogate, and Vincent Ng

Tuesday, November 1 – 09:00–12:30

Room 616B

In the early days of the statistical NLP era, many language processing tasks were tackled using the so-called pipeline architecture: the given task is broken into a series of sub-tasks such that the output of one sub-task is an input to the next sub-task in the sequence. The pipeline architecture is appealing for various reasons, including modularity, modeling convenience, and manageable computational complexity. However, it suffers from the error propagation problem: errors made in one sub-task are propagated to the next sub-task in the sequence, leading to poor accuracy on that sub-task, which in turn leads to more errors downstream. Another disadvantage associated with it is lack of feedback: errors made in a sub-task are often not corrected using knowledge uncovered while solving another sub-task down the pipeline.

Realizing these weaknesses, researchers have turned to joint inference approaches in recent years. One such approach involves the use of Markov logic, which is defined as a set of weighted first-order logic formulas and, at a high level, unifies first-order logic with probabilistic graphical models. It is an ideal modeling language (knowledge representation) for compactly representing relational and uncertain knowledge in NLP. In a typical use case of MLNs in NLP, the application designer describes the background knowledge using a few first-order logic sentences and then uses software packages such as Alchemy, Tuffy , and Markov the beast to perform learning and inference (prediction) over the MLN. However,

---

Deepak Venugopal completed his Ph.D. at the University of Texas at Dallas in 2015 after which he joined the department of computer science at the University of Memphis where he is currently an assistant professor. His main research interests are in Probabilistic Graphical Models, Statistical Relational Models and their applications in natural language processing and cyber-security. His research work has resulted in several key techniques for scalable inference and learning in Statistical Relational Models.

Vibhav Gogate is an Assistant Professor in the Computer Science Department at the University of Texas at Dallas. He got his Ph.D. at University of California, Irvine in 2009 and then did a two-year post-doc at University of Washington. His current research focus is on probabilistic graphical models and its first-order logic based extensions such as Markov logic. He is best known for his work on inference algorithms that combine the power of logic and probability including lifted probabilistic inference algorithms. He is the co-winner of the last two probabilistic inference competitions the 2010 UAI approximate inference challenge and the 2012 PASCAL probabilistic inference competition.

Vincent Ng is an Associate Professor of Computer Science and a member of the Human Language Technology Research Institute at the University of Texas at Dallas. He is best known for his work on coreference resolution in the NLP community. He has recently been collaborating with the co-authors of this proposal on modeling complex NLP tasks using Markov Logic.

despite its obvious advantages, over the years, researchers and practitioners have found it difficult to use MLNs effectively in many NLP applications. The main reason for this is that it is hard to scale inference and learning algorithms for MLNs to large datasets and complex models, that are typical in NLP.

In this tutorial, we will introduce the audience to recent advances in scaling up inference and learning in MLNs as well as new approaches to make MLNs a "black-box" for NLP applications (with only minor tuning required on the part of the user). Specifically, we will introduce attendees to a key idea that has emerged in the MLN research community over the last few years, lifted inference, which refers to inference techniques that take advantage of symmetries (e.g., synonyms), both exact and approximate, in the MLN. We will describe how these next-generation inference techniques can be used to perform effective joint inference. We will also present our new software package for inference and learning in MLNs, Alchemy 2.0, which is based on lifted inference, focusing primarily on how it can be used to scale up inference and learning in large models and datasets for applications such as semantic similarity determination, information extraction and question answering.

Lifelong Machine Learning for Natural Language Processing

Zhiyuan Chen and Bing Liu

Tuesday, November 1 – 09:00–12:30

Room 616A

Machine learning (ML) has been successfully used as a prevalent approach to solving numerous NLP problems. However, the classic ML paradigm learns in isolation. That is, given a dataset, an ML algorithm is executed on the dataset to produce a model without using any related or prior knowledge. Although this type of isolated learning is very useful, it also has serious limitations as it does not accumulate knowledge learned in the past and use the knowledge to help future learning, which is the hallmark of human learning and human intelligence. Lifelong machine learning (LML) aims to achieve this capability. Specifically, it aims to design and develop computational learning systems and algorithms that learn as humans do, i.e., retaining the results learned in the past, abstracting knowledge from them, and using the knowledge to help future learning. In this tutorial, we will introduce the existing research of LML and to show that LML is very suitable for NLP tasks and has potential to help NLP make major progresses.

Zhiyuan Chen, Google Zhiyuan Chen completed his Ph.D. at the University of Illinois at Chicago (UIC) and joined Google in 2016. His research interests include Text Mining, Machine Learning, and Statistical Natural Language Processing. He has designed several lifelong learning algorithms to automatically mine valuable information from text documents. He has published more than 15 full research papers at premier conferences.

Bing Liu, University of Illinois at Chicago (UIC) Bing Liu is a professor of Computer Science at the University of Illinois at Chicago (UIC). He received his PhD in Artificial Intelligence from the University of Edinburgh. His current research interests include sentiment analysis, NLP, data mining, and machine learning. He has published extensively in top conferences and journals, and also three books. Two of his papers received test-of-time awards from KDD. He is a Fellow of ACM, AAAI, and IEEE.

Neural Networks for Sentiment Analysis

Yue Zhang and Duy Tin Vo

Tuesday, November 1 – 14:00–17:30

Room 616A

Sentiment analysis has been a major research topic in natural language processing (NLP). Traditionally, the problem has been attacked using discrete models and manually-defined sparse features. Over the past few years, neural network models have received increased research efforts in most sub areas of sentiment analysis, giving highly promising results. A main reason is the capability of neural models to automatically learn dense features that capture subtle semantic information over words, sentences and documents, which are difficult to model using traditional discrete features based on words and ngram patterns. This tutorial gives an introduction to neural network models for sentiment analysis, discussing the mathematics of word embeddings, sequence models and tree structured models and their use in sentiment analysis on the word, sentence and document levels, and fine-grained sentiment analysis. The tutorial covers a range of neural network models (e.g. CNN, RNN, RecNN, LSTM) and their extensions, which are employed in four main subtasks of sentiment analysis:

- Sentiment-oriented embeddings;
- Sentence-level sentiment;
- Document-level sentiment;
- Fine-grained sentiment.

The content of the tutorial is divided into 3 sections of 1 hour each. We assume that the audience is familiar with linear algebra and basic neural network structures, introduce the mathematical details of the most typical models. First, we will introduce the sentiment

---

Yue Zhang, Assistant Professor, Singapore University of Technology and Design Yue Zhang is an assistant professor at Singapore University of Technology and Design (SUTD). Before joining SUTD, he worked as a postdoctoral research fellow at University of Cambridge. He received his DPhil and MSc degrees from University of Oxford, and BEng from Tsinghua University. His main research interest includes machine learning, syntactic parsing and text mining. Over the recent years, he worked intensively on deep learning and sentiment analysis. He serves in the standing review committee of TACL, and was area chair of COLING'14, NAACL'15 and EMNLP'15. Yue Zhang has given a tutorial at NAACL'10 and a tutorial at ACL'14.

Duy Tin Vo, PhD Candidate, Singapore University of Technology and Design Duy Tin Vo is a PhD candidate at Singapore University of Technology and Design (SUTD) under the supervision of Yue Zhang. He received his BEng from Cantho University. His current research focuses on applying machine learning and deep learning in sentiment analysis.

analysis task, basic concepts related to neural network models for sentiment analysis, and show detail approaches to integrate sentiment information into embeddings. Sentence-level models will be described in the second section. Finally, we will discuss neural network models use for document-level and fine-grained sentiment.

Continuous Vector Spaces for Cross-language NLP Applications

Rafael E. Banchs

Tuesday, November 1 – 14:00–17:30

Room 616B

The mathematical metaphor offered by the geometric concept of distance in vector spaces with respect to semantics and meaning has been proven to be useful in many monolingual natural language processing applications. There is also some recent and strong evidence that this paradigm can also be useful in the cross-language setting. In this tutorial, we present and discuss some of the most recent advances on exploiting the vector space model paradigm in specific cross-language natural language processing applications, along with a comprehensive review of the theoretical background behind them.

First, the tutorial introduces some fundamental concepts of distributional semantics and vector space models. More specifically, the concepts of distributional hypothesis and term-document matrices are revised, followed by a brief discussion on linear and non-linear dimensionality reduction techniques and their implications to the parallel distributed approach to semantic cognition. Next, some classical examples of using vector space models in monolingual natural language processing applications are presented. Specific examples in the areas of information retrieval, related term identification and semantic compositionality are described.

Then, the tutorial focuses its attention on the use of the vector space model paradigm in cross-language applications. To this end, some recent examples are presented and discussed in detail, addressing the specific problems of cross-language information retrieval, cross-language sentence matching, and machine translation. Some of the most recent developments in the area of Neural Machine Translation are also discussed.

Finally, the tutorial concludes with a discussion about current and future research problems related to the use of vector space models in cross-language settings. Future avenues for scientific research are described, with major emphasis on the extension from vector and matrix representations to tensors, as well as the problem of encoding word position information into the vector-based representations.

Rafael E Banchs, Researcher, Institute for Infocomm Research, Singapore Rafael is a Research Scientist at the Institute for Infocomm Research, in Singapore, where he leads the Dialogue Technology Lab of the Human Language Technology Department. His main area of research is focused on the construction and use of semantic representations to support different natural language processing applications, including machine translation, information retrieval, natural language understanding and chat-oriented dialogue.

Methods and Theories for Large-scale Structured Prediction

Xu Sun and Yansong Feng

Tuesday, November 1 – 14:00–17:30

Salon K

Many important NLP tasks are casted as structured prediction problems, and try to predict certain forms of structured output from the input. Examples of structured prediction include POS tagging, named entity recognition, PCFG parsing, dependency parsing, machine translation, and many others. When apply structured prediction to a specific NLP task, there are the following challenges:

1. Model selection: Among various models/algorithms with different characteristics, which one should we choose for a specific NLP task?

2. Training: How to train the model parameters effectively and efficiently?

3. Overfitting: To achieve good accuracy on test data, it is important to control the overfitting from the training data. How to control the overfitting risk for structured prediction?

This tutorial will provide a clear overview of recent advances in structured prediction methods and theories, and address the above issues when we apply structured prediction to NLP tasks. We will introduce large margin methods (e.g., perceptrons, MIRA), graphical models (e.g., CRFs), and deep learning methods (e.g., RNN, LSTM), and show the respective advantages and disadvantages for NLP applications. For the training algorithms, we will introduce online/ stochastic training methods, and we will introduce parallel online/stochastic

---

Xu Sun, Associate Professor, Peking University Xu Sun is currently an Associate Professor at the School of EECS, Peking University. He got PhD from The University of Tokyo in 2010. He worked at The University of Tokyo, Cornell University, and The Hong Kong Polytechnic University as research fellow/associate from 2010 to 2012. His research interests focus on natural language processing and machine learning, especially on structured natural language processing and structured prediction, and with related publications on ACL, NIPS, IJCAI, Comput. Linguist., IEEE T-KDE, etc. He has been area chair of EMNLP; program committee member of ACL, IJCAI, AAAI, NAACL; journal reviewer of IEEE T-PAMI, Comput. Linguist., IEEE T-ASL, and so on.

Yansong Feng, Lecturer, Peking University Yansong Feng is a Lecturer at the Institute of Computer Science and Technology, Peking University. He received his Ph.D. degree from ILCC, the University of Edinburgh and worked as a postdoctoral research assistant there in 2011. His research interests focus on applying machine learning techniques to natural language processing and information retrieval applications, such as information extraction, question answering, knowledge acquisition, and so on. His works have been published in top refereed international journals and conferences, such as IEEE TPAMI, ACL, VLDB, EMNLP, NAACL, AAAI, etc. He has served as program committee member of AAAI, EMNLP, NAACL, WWW; journal reviewer of IEEE TASLP, IEEE TKDE, TACL, and so on.

learning algorithms and theories to speed up the training (e.g., the Hogwild algorithm). For controlling the overfitting from training data, we will introduce the weight regularization methods, structure regularization, and implicit regularization methods.

3 Workshops

Tuesday

Saturday

Workshop 1: NLPSM
4th International Workshop on Natural Language Processing for Social Media

Organizers: *Jane Yung-jen Hsu, Lun-Wei Ku, and Cheng-Te Li*

Venue: Salon J

Tuesday, November 1, 2016

08:50–09:00 **Opening**

09:00–10:00 **Keynote Speech: Cristian Danescu-Niculescu-Mizil**

10:00–10:30 **Short Paper Session I**

10:00–10:15 Identifying and Categorizing Disaster-Related Tweets
Kevin Stowe, Michael J. Paul, Martha Palmer, Leysia Palen, and Kenneth Anderson

10:15–10:30 Identifying Eyewitness News-worthy Events on Twitter
Erika Doggett and Alejandro Cantarero

10:30–11:00 **Coffee Break**

11:00–12:15 **Short Paper Session II**

11:00–11:15 Why Do They Leave: Modeling Participation in Online Depression Forums
Farig Sadeque, Ted Pedersen, Thamar Solorio, Prasha Shrestha, Nicolas Rey-Villamizar, and Steven Bethard

11:15–11:30 Twitter at the Grammys: A Social Media Corpus for Entity Linking and Disambiguation
Mark Dredze, Nicholas Andrews, and Jay DeYoung

11:30–11:45 Steps Toward Automatic Understanding of the Function of Affective Language in Support Groups
Amit Navindgi, Caroline Brun, Cécile Boulard Masson, and Scott Nowson

11:45–12:00 Detecting Social Roles in Twitter
Sunghwan Mac Kim, Stephen Wan, and Cecile Paris

12:00–12:15 Identifying Sensible Participants in Online Discussions
Siddharth Jain

12:30–14:00 **Lunch**

14:00–15:30 **Long Paper Session I**

14:00–14:30 emoji2vec: Learning Emoji Representations from their Description
Ben Eisner, Tim Rocktäschel, Isabelle Augenstein, Matko Bosnjak, and Sebastian Riedel

14:30–15:00 Learning Latent Local Conversation Modes for Predicting Comment Endorsement in Online Discussions
Hao Fang, Hao Cheng, and Mari Ostendorf

15:00–15:30 Witness Identification in Twitter
Rui Fang, Armineh Nourbakhsh, XIAOMO LIU, Sameena Shah, and Quanzhi Li

15:30–16:00 **Coffee Break / Poster Session and Discussion**

16:00–17:30 **Long Paper Session II**

16:00–16:30 How Do I Look? Publicity Mining From Distributed Keyword Representation of Socially Infused News Articles
Yu-Lun Hsieh, Yung-Chun Chang, Chun-Han Chu, and Wen-Lian Hsu

16:30–17:00 Hierarchical Character-Word Models for Language Identification
Aaron Jaech, George Mulcaire, Shobhit Hathi, Mari Ostendorf, and Noah A. Smith

17:00–17:30 Human versus Machine Attention in Document Classification: A Dataset with Crowdsourced Annotations
Nikolaos Pappas and Andrei Popescu-Belis

17:30–17:40 **Award and Closing**

Workshop 2: LCS
Computational Approaches to Linguistic Code Switching

Organizers: *Mona Diab, Pascale Fung, Mahmoud Ghoneim, Julia Hirschberg, and Thamar Solorio*

Venue: Room 417 AB

Tuesday, November 1, 2016

08:45–10:30 **Session 1: Opening Session**

08:45–09:00 *Welcome Remarks*

09:00–10:00 *Keynote Speakers*

10:00–10:30 Challenges of Computational Processing of Code-Switching
Özlem Çetinoğlu, Sarah Schulz, and Ngoc Thang Vu

10:30–11:00 **Coffee Break**

11:00–12:30 **Session 2: Workshop Talks**

11:00–11:30 Simple Tools for Exploring Variation in Code-switching for Linguists
Gualberto A. Guzman, Jacqueline Serigos, Barbara E. Bullock, and Almeida Jacqueline Toribio

11:30–12:00 Word-Level Language Identification and Predicting Codeswitching Points in Swahili-English Language Data
Mario Piergallini, Rouzbeh Shirvani, Gauri S. Gautam, and Mohamed Chouikha

12:00–12:30 Part-of-speech Tagging of Code-mixed Social Media Content: Pipeline, Stacking and Joint Modelling
Utsab Barman, Joachim Wagner, and Jennifer Foster

12:30–14:00 **Lunch**

14:00–15:30 **Session 3: Shared Task**

14:00–14:30 Overview for the Second Shared Task on Language Identification in Code-Switched Data
Giovanni Molina, Fahad AlGhamdi, Mahmoud Ghoneim, Abdelati Hawwari, Nicolas Rey-Villamizar, Mona Diab, and Thamar Solorio

14:30–15:00 Multilingual Code-switching Identification via LSTM Recurrent Neural Networks
Younes Samih, Suraj Maharjan, Mohammed Attia, Laura Kallmeyer, and Thamar Solorio

15:00–15:30 A Neural Model for Language Identification in Code-Switched Tweets
Aaron Jaech, George Mulcaire, Mari Ostendorf, and Noah A. Smith

15:30–16:00 **Coffee Break**

16:00–18:00 **Session 4: Panel Discussion and Poster Sessions**

16:00–16:45 *Panel Discussion*

16:45–18:00 *Poster Session*

- SAWT: Sequence Annotation Web Tool
  Younes Samih, Wolfgang Maier, and Laura Kallmeyer
- Accurate Pinyin-English Codeswitched Language Identification
  Meng Xuan Xia and Jackie Chi Kit Cheung
- Unraveling the English-Bengali Code-Mixing Phenomenon
  Arunavha Chanda, Dipankar Das, and Chandan Mazumdar
- Part-of-speech Tagging of Code-Mixed Social Media Text
  Souvick Ghosh, Satanu Ghosh, and Dipankar Das
- Part of Speech Tagging for Code Switched Data
  Fahad AlGhamdi, Giovanni Molina, Mona Diab, Thamar Solorio, Abdelati Hawwari, Victor Soto, and Julia Hirschberg
- The George Washington University System for the Code-Switching Workshop Shared Task 2016
  Mohamed Al-Badrashiny and Mona Diab
- Columbia-Jadavpur submission for EMNLP 2016 Code-Switching Workshop Shared Task: System description
  Arunavha Chanda, Dipankar Das, and Chandan Mazumdar
- The Howard University System Submission for the Shared Task in Language Identification in Spanish-English Codeswitching
  Rouzbeh Shirvani, Mario Piergallini, Gauri Shankar Gautam, and Mohamed Chouikha
- Codeswitching Detection via Lexical Features in Conditional Random Fields
  Prajwol Shrestha
- Language Identification in Code-Switched Text Using Conditional Random Fields and Babelnet
  Utpal Kumar Sikdar and Björn Gambäck
- Codeswitching language identification using Subword Information Enriched Word Vectors
  Meng Xuan Xia

Workshop 3: SPNLP
Structured Prediction for NLP

Organizers: *Kai-Wei Chang, Alexander Rush, Vivek Srikumar, and Ming-Wei Chang*

Venue: The Gallery

Saturday, November 5, 2016

09:00–10:30 **Section 1**

09:00–09:15 *Welcome (Organizers)*

09:15–10:00 *Invited Talk (Kristina Toutanova)*

10:00–10:30 Inside-Outside and Forward-Backward Algorithms Are Just Backprop (tutorial paper)
Jason Eisner

10:30–11:00 **Coffee break**

11:00–12:30 **Section 2**

11:00–11:45 *Invited Talk (Andrew McCallum)*

11:45–12:30 *Panel (Panel: Moderator - Hal Daume)*

12:30–14:00 **Lunch**

14:00–15:30 **Section 3**

- Research on attention memory networks as a model for learning natural language inference
  Degen Huang
- A Joint Model of Rhetorical Discourse Structure and Summarization
  Naman Goyal and Jacob Eisenstein
- Posterior regularization for Joint Modeling of Multiple Structured Prediction Tasks with Soft Constraints
  Kartik Goyal and Chris Dyer
- A Study of Imitation Learning Methods for Semantic Role Labeling
  Travis Wolfe, Mark Dredze, and Benjamin Van Durme
- Introducing DRAIL – a Step Towards Declarative Deep Relational Learning
  Xiao Zhang, María Leonor Pacheco, Chang Li, and Dan Goldwasser

14:45–15:30 *Invited Talk (Raquel Urtasun)*

14:00–14:45 **Poster**

15:30–16:00 **Coffee Break**

16:00–17:50 **Section 4**

16:00–16:45 *Invited Talk (Dzmitry Bahdanau)*

16:45–17:15 Unsupervised Neural Hidden Markov Models
Ke M. Tran, Yonatan Bisk, Ashish Vaswani, Daniel Marcu, and Kevin Knight

17:15–17:30 **Closing**

Workshop 4: CSS
Natural Language Processing and Computational Social Science

Organizers: *Dirk Hovy, David Bamman, A. Seza Dogruoz, Jacob Eisenstein, David Jurgens, Brendan O'Connor, Alice Oh, Oren Tsur, and Svitlana Volkova*

Venue: Broken Spoke

Saturday, November 5, 2016

09:00–10:30 **Session 1**

09:00–09:15 *Welcome (Workshop organizers)*

09:15–10:00 *Invited talk (James Pennebaker)*

10:00–10:30 **Doctoral consortium presentations**

10:30–11:00 **coffee break**

11:00–12:30 **Session 2**

11:00–11:45 *Invited talk (Hanna Wallach)*

11:45–12:30 *Invited talk (Cristian Danescu-Niculescu-Mizil)*

12:30–14:00 **Lunch break**

14:00–15:30 **Session 3**

14:00–14:45 *Invited talk (Jason Baldridge)*

14:45–15:30 **1-minute poster madness**

- Relating semantic similarity and semantic association to how humans label other people
  Kenneth Joseph and Kathleen M. Carley
- Identifying News from Tweets
  Jesse Freitas and Heng Ji
- Obfuscating Gender in Social Media Writing
  Sravana Reddy and Kevin Knight
- Social Proof: The Impact of Author Traits on Influence Detection
  Sara Rosenthal and Kathy McKeown
- Generating Politically-Relevant Event Data
  John Beieler
- User profiling with geo-located posts and demographic data
  Adam Poulston, Mark Stevenson, and Kalina Bontcheva
- Gov2Vec: Learning Distributed Representations of Institutions and Their Legal Text
  John J. Nay

- #WhoAmI in 160 Characters? Classifying Social Identities Based on Twitter Profile Descriptions
  Anna Priante, Djoerd Hiemstra, Tijs van den Broek, Aaqib Saeed, Michel Ehrenhard, and Ariana Need
- Identifying Stance by Analyzing Political Discourse on Twitter
  Kristen Johnson and Dan Goldwasser
- Learning Linguistic Descriptors of User Roles in Online Communities
  Alex Wang, William L. Hamilton, and Jure Leskovec
- The Effects of Data Collection Methods in Twitter
  Sunghwan Mac Kim, Stephen Wan, Cecile Paris, Jin Brian, and Bella Robinson
- Expressions of Anxiety in Political Texts
  Ludovic Rheault
- Constructing an Annotated Corpus for Protest Event Mining
  Peter Makarov, Jasmine Lorenzini, and Hanspeter Kriesi
- Demographer: Extremely Simple Name Demographics
  Rebecca Knowles, Josh Carroll, and Mark Dredze
- Bag of What? Simple Noun Phrase Extraction for Text Analysis
  Abram Handler, Matthew Denny, Hanna Wallach, and Brendan O'Connor
- News Sentiment and Cross-Country Fluctuations
  Samuel Fraiberger
- The Clinical Panel: Leveraging Psychological Expertise During NLP Research
  Glen Coppersmith, Kristy Hollingshead, H. Andrew Schwartz, Molly Ireland, Rebecca Resnik, Kate Loveys, April Foreman, and Loring Ingraham
- Are You a Racist or Am I Seeing Things? Annotator Influence on Hate Speech Detection on Twitter
  Zeerak Waseem
- Disentangling Topic Models: A Cross-cultural Analysis of Personal Values through Words
  Steven Wilson, Rada Mihalcea, Ryan Boyd, and James Pennebaker

15:30–16:00 **coffee break**

16:00–16:45 **posters**

16:00–17:30 **Session 4**

16:45–17:30 *Invited talk (Molly Roberts)*

Workshop 5: LOUHI
Louhi 2016: the Seventh International Workshop on Health Text Mining and Information Analysis

Organizers: *Cyril Grouin, Thierry Hamon, Aurélie Névéol, and Pierre Zweigenbaum*

Venue: TBD

Saturday, November 5, 2016

09:00–10:15 **Session I - Machine-Learning**

09:00–09:25 An Investigation of Recurrent Neural Architectures for Drug Name Recognition
Raghavendra Chalapathy, Ehsan Zare Borzeshi, and Massimo Piccardi

09:25–09:50 Clinical Text Prediction with Numerically Grounded Conditional Language Models
Georgios Spithourakis, Steffen Petersen, and Sebastian Riedel

09:50–10:15 Modelling Radiological Language with Bidirectional Long Short-Term Memory Networks
Savelie Cornegruta, Robert Bakewell, Samuel Withey, and Giovanni Montana

10:15–10:30 **Session II - Boosters**

11:00–12:30 **Session III - Posters**

- Data Resource Acquisition from People at Various Stages of Cognitive Decline — Design and Exploration Considerations
Dimitrios Kokkinakis, Kristina Lundholm Fors, and Arto Nordlund
- Analysis of Anxious Word Usage on Online Health Forums
Nicolas Rey-Villamizar, Prasha Shrestha, Farig Sadeque, Steven Bethard, Ted Pedersen, Arjun Mukherjee, and Thamar Solorio
- Retrofitting Word Vectors of MeSH Terms to Improve Semantic Similarity Measures
Zhiguo Yu, Trevor Cohen, Byron Wallace, Elmer Bernstam, and Todd Johnson
- Unsupervised Resolution of Acronyms and Abbreviations in Nursing Notes Using Document-Level Context Models
Katrin Kirchhoff and Anne M. Turner
- Low-resource OCR error detection and correction in French Clinical Texts
Eva D'hondt, Cyril Grouin, and Brigitte Grau
- Citation Analysis with Neural Attention Models
Tsendsuren Munkhdalai, John Lalor, and Hong Yu
- Replicability of Research in Biomedical Natural Language Processing: a pilot evaluation for a coding task
Aurelie Neveol, Kevin Cohen, Cyril Grouin, and Aude Robert

12:30–14:00 **Lunch break**

14:00–15:30 **Session IV - Invited talk**

14:00–15:30 NLP and Online Health Reports: What do we say and what do we mean?
Nigel Collier

16:00–17:15 **Session V - NLP for literature and clinical documents**

16:00–16:25 Leveraging coreference to identify arms in medical abstracts: An experimental study
Elisa Ferracane, Iain Marshall, Byron C. Wallace, and Katrin Erk

16:25–16:50 Hybrid methods for ICD-10 coding of death certificates
Pierre Zweigenbaum and Thomas Lavergne

16:50–17:15 Exploring Query Expansion for Entity Searches in PubMed
Chung-Chi Huang and Zhiyong Lu

Workshop 6: CNS
Computing News Storylines

Organizers: *Tommaso Caselli, Ben Miller, Marieke van Erp, Piek Vossen, and David Caswell*

Venue: Moody

Friday, November 5, 2016

09:00–10:30 **Session 1:**

09:10–10:10 *Bridging the Gap between Event Macro-structures and Event Micro-structures (Ed Hovy, Language Technologies Institute, CMU)*

09:00–09:10 **Welcome and Opening Remarks**

10:30–11:00 **Coffee Break**

11:00–12:30 **Session 2:**

11:00–11:25 Computable News Ecosystems: Roles for Humans and Machines
David Caswell

11:25–11:50 Storyline detection and tracking using Dynamic Latent Dirichlet Allocation
Daniel Bruggermann, Yannik Hermey, Carsten Orth, Darius Schneider, Stefan Selzer, and Gerasimos Spanakis

11:50–12:15 Real-time News Story Detection and Tracking with Hashtags
Gevorg Poghosyan and Georgiana Ifrim

12:15–12:30 Nonparametric Bayesian Storyline Detection from Microtexts
Vinodh Krishnan and Jacob Eisenstein

12:30–14:00 **Lunch**

14:00–15:30 **Session 3:**

14:00–14:25 Automatic Identification of Narrative Diegesis and Point of View
Joshua Eisenberg and Mark Finlayson

14:25–14:50 Richer Event Description: Integrating event coreference with temporal, causal and bridging annotation
Tim O'Gorman, Kristin Wright-Bettner, and Martha Palmer

14:50–15:15 NASTEA: Investigating Narrative Schemas through Annotated Entities
Dan Simonson and Anthony Davis

15:15–15:30 The Storyline Annotation and Representation Scheme (StaR): A Proposal
Tommaso Caselli and Piek Vossen

16:00–17:00 **Discussion and Closing Remarks**

16:00–17:00 **Session 4:**

Workshop 7: UBLP
Uphill Battles in Language Processing: Scaling early achievements to robust methods

Organizers: *Annie Louis, Michael Roth, Bonnie Webber, Michael White, and Luke Zettlemoyer*

Venue: Continental Ballroom II

Saturday, November 5, 2016

09:00–10:30 **Text Understanding**

09:00–10:20 *Invited talks, followed by discussion (Hal Daume III, Andrew Kehler, Chris Manning, Marie-Catherine de Marneffe)*

10:20–10:30 *Poster Boasters*

10:20–10:20 An Analysis of Prerequisite Skills for Reading Comprehension
Saku Sugawara and Akiko Aizawa

10:20–10:20 Bridging the gap between computable and expressive event representations in Social Media
Darina Benikova and Torsten Zesch

10:20–10:20 Statistical Script Learning with Recurrent Neural Networks
Karl Pichotta and Raymond Mooney

10:20–10:20 Moving away from semantic overfitting in disambiguation datasets
Marten Postma, Filip Ilievski, Piek Vossen, and Marieke van Erp

10:20–10:20 Unsupervised Event Coreference for Abstract Words
Dheeraj Rajagopal, Eduard Hovy, and Teruko Mitamura

10:20–10:20 Towards Broad-coverage Meaning Representation: The Case of Comparison Structures
Omid Bakhshandeh and James Allen

10:30–11:00 **Coffee break**

11:00–12:30 **Natural Language Generation**

11:00–12:20 *Invited talks, followed by discussion (Ioannis Konstas, Kathleen McKeown, Margaret Mitchell, Donia Scott)*

12:20–12:30 *Poster Boasters*

12:20–12:20 DialPort: A General Framework for Aggregating Dialog Systems
Tiancheng Zhao, Kyusong Lee, and Maxine Eskenazi

12:20–12:20 C2D2E2: Using Call Centers to Motivate the Use of Dialog and Diarization in Entity Extraction
Ken Church, Weizhong Zhu, and Jason Pelecanos

12:20–12:20 Visualizing the Content of a Children's Story in a Virtual World: Lessons Learned
Quynh Ngoc Thi Do, Steven Bethard, and Marie-Francine Moens

12:20–12:20 Stylistic Transfer in Natural Language Generation Systems Using Recurrent Neural Networks
Jad Kabbara and Jackie Chi Kit Cheung

12:20–12:20 Using Language Groundings for Context-Sensitive Text Prediction
Timothy Lewis, Cynthia Matuszek, Amy Hurst, and Matthew Taylor

12:20–12:20 Towards a continuous modeling of natural language domains
Sebastian Ruder, Parsa Ghaffari, and John G. Breslin

12:30–14:00 **Lunch break**

14:00–15:20 **Dialogue and Speech**

14:00–15:20 *Invited talks, followed by discussion (David DeVault, Mark Liberman, Diane Litman, Amanda Stent)*

15:20–16:00 **Coffee break + poster session**

16:00–16:30 **Poster session (continued)**

16:30–17:50 **Grounded Language**

16:30–17:50 *Invited talks, followed by discussion (James Allen, Joyce Chai, Yejin Choi, Mark Steedman)*

4

Main Conference: Wednesday, November 2

Overview

07:30 – 17:30 **Registration Day 1** *Salon GHJ Prefunction*

08:00 – 08:40 **Morning Coffee** *Salon GHJ Prefunction*

08:40 – 10:00 **Session P1: Plenary Session I** *Salon FG*

08:40 – 09:00 *Opening Remarks (General Chair)*

09:00 – 10:00 *Invited Talk: Learning in extended and approximate Rational Speech Acts models (Christopher Potts)*

10:00 – 10:30 **Coffee Break** *Salon GHJ Prefunction*

10:30 – 12:10 **Session 1**

Parsing and Syntax *Salon FG*	Information Extraction *Salon J*	Psycholinguistics / Machine Learning *Room 616 AB*

12:10 – 13:40 **Lunch**

13:40 – 15:20 **Session 2**

Reading Comprehension and Question Answering *Salon FG*	Embeddings of Linguistic Structure *Salon J*	Sentiment and Opinion Analysis *Room 616 AB*

15:20 – 15:50 **Coffee Break** *Salon GHJ Prefunction*

15:50 – 17:30 **Session 3**

Neural Machine Translation *Salon FG*	Semi-supervised and Minimally Supervised Learning *Salon J*	Summarization and Generation *Room 616 AB*

17:30 – 17:45 **Break**

17:45 – 18:15 **Poster Spotlight A (Half-minute Madness)** *Salon FG*

18:15 – 20:15 **Poster Session A** *Salon H and J*

Invited Speaker: Christopher Potts

Learning in extended and approximate Rational Speech Acts models

Wednesday, November 2, 2016,

Abstract: The Rational Speech Acts (RSA) model treats language use as a recursive process in which probabilistic speaker and listener agents reason about each other's intentions to enrich, and negotiate, the semantics of their language along broadly Gricean lines. RSA builds on early work by the philosopher David Lewis and others on signaling systems as well as more recent developments in Bayesian cognitive modeling. Over the last five years, RSA has been shown to provide a unified account of numerous core phenomena in pragmatics, including metaphor, hyperbole, sarcasm, politeness, and a wide range of conversational implicatures. Its precise, quantitative nature has also facilitated an outpouring of new experimental work on these phenomena. However, applications of RSA to large-scale problems in NLP and AI have so far been limited, because the exact version of the model is intractable along several dimensions. In this talk, I'll report on recent progress in approximating RSA in ways that retains its core properties while enabling application to large datasets and complex environments in which language and action are brought together.

Biography: Christopher Potts is Professor of Linguistics and, by courtesy, of Computer Science, at Stanford, and Director of the Center for the Study of Language and Information (CSLI) at Stanford. He earned his BA in Linguistics from NYU in 1999 and his PhD from UC Santa Cruz in 2003. He was on the faculty in Linguistics at UMass Amherst from 2003 until 2009, when he headed west once again, to join Stanford Linguistics. He was a co-editor at Linguistic Inquiry 2004–2006, an associate editor at Linguistics and Philosophy 2009–2012, and has been an Action Editor at TACL since 2014. In his research, he uses computational methods to explore how emotion is expressed in language and how linguistic production and interpretation are influenced by the context of utterance. He is the author of the 2005 book The Logic of Conventional Implicatures as well as numerous scholarly papers in computational and theoretical linguistics.

Session 1 Overview

Track A	**Track B**	**Track C**	
Parsing and Syntax Salon FG	*Information Extraction* Salon J	*Psycholinguistics / Machine Learning* Room 616 AB	
Span-Based Constituency Parsing with a Structure-Label System and Provably Optimal Dynamic Oracles *Cross and Huang*	Distinguishing Past, On-going, and Future Events: The EventStatus Corpus *Huang, Cases, Jurafsky, Condoravdi, and Riloff*	Modeling Human Reading with Neural Attention *Hahn and Keller*	10:30
Rule Extraction for Tree-to-Tree Transducers by Cost Minimization *Martínez-Gómez and Miyao*	Nested Propositions in Open Information Extraction *Bhutani, Jagadish, and Radev*	Comparing Computational Cognitive Models of Generalization in a Language Acquisition Task *Barak, Goldberg, and Stevenson*	10:55
A Neural Network for Coordination Boundary Prediction *Ficler and Goldberg*	A Position Encoding Convolutional Neural Network Based on Dependency Tree for Relation Classification *Yang, Tong, Ma, and Deng*	Rationalizing Neural Predictions *Lei, Barzilay, and Jaakkola*	11:20
Using Left-corner Parsing to Encode Universal Structural Constraints in Grammar Induction *Noji, Miyao, and Johnson*	Learning to Recognize Discontiguous Entities *Muis and Lu*	Deep Multi-Task Learning with Shared Memory for Text Classification *Liu, Qiu, and Huang*	11:45

Session 1A: Parsing and Syntax

Salon FG

Span-Based Constituency Parsing with a Structure-Label System and Provably Optimal Dynamic Oracles

James Cross and Liang Huang 10:30–10:55

Parsing accuracy using extremely efficient greedy transition-based parsers has improved dramatically in recent years thanks to neural-network learning models. Despite striking results in dependency parsing, however, neural models have not surpassed state-of-the-art approaches in constituency parsing. To remedy this, we introduce a new parsing system which uses a stack of sentence spans, represented by a bare minimum of LSTM features. We also describe a dynamic oracle for this constituency parsing system. Training with this oracle, we achieve the best performance on Penn Treebank of any parser that does not use reranking or any external data.

Rule Extraction for Tree-to-Tree Transducers by Cost Minimization

Pascual Martínez-Gómez and Yusuke Miyao 10:55–11:20

Tree transducers that model expressive linguistic phenomena often require word-alignments and a heuristic rule extractor to induce their grammars. However, when the corpus of tree/string pairs is small compared to the size of the vocabulary or the complexity of the grammar, word-alignments are unreliable. We propose a general rule extraction algorithm that uses cost functions over tree fragments, and formulate the extraction as a cost minimization problem. As a by-product, we are able to introduce back-off states at which some cost functions generate right-hand-sides of previously unseen left-hand-sides, thus creating transducer rules "on-the-fly". We test the generalization power of our induced tree transducers on a QA task over a large Knowledge Base, obtaining a reasonable syntactic accuracy and effectively overcoming the typical lack of rule coverage.

A Neural Network for Coordination Boundary Prediction

Jessica Ficler and Yoav Goldberg 11:20–11:45

We propose a neural-network based model for coordination boundary prediction. The network is designed to incorporate two signals: the similarity between conjuncts and the observation that replacing the whole coordination phrase with a conjunct tends to produce a coherent sentences. The modeling makes use of several LSTM networks. The model is trained solely on conjunction annotations in treebank, without using external resources. We show improvements on predicting coordination boundaries on the PTB compared to two state-of-the-art parsers; as well as improvement over previous coordination boundary prediction systems on the Genia corpus.

Using Left-corner Parsing to Encode Universal Structural Constraints in Grammar Induction

Hiroshi Noji, Yusuke Miyao, and Mark Johnson 11:45–12:10

Center-embedding is difficult to process and is known as a rare syntactic construction across languages. In this paper we describe a method to incorporate this assumption into the grammar induction tasks by restricting the search space of a model to trees with limited center-embedding. The key idea is the tabulation of *left-corner* parsing, which captures the degree of center-embedding of a parse via its stack depth. We apply the technique to learning of famous generative model, the dependency model with valence (Klein and Manning, 2004). Cross-linguistic experiments on Universal Dependencies show that often our method boosts the performance from the baseline, and competes with the current state-of-the-art model in a number of languages.

Session 1B: Information Extraction

Salon J

Distinguishing Past, On-going, and Future Events: The EventStatus Corpus

Ruihong Huang, Ignacio Cases, Dan Jurafsky, Cleo Condoravdi, and Ellen Riloff 10:30–10:55

Determining whether a major societal event has already happened, is still on-going, or may occur in the future is crucial for event prediction, timeline generation, and news summarization. We introduce a new task and a new corpus, EventStatus, which has 4500 English and Spanish articles about civil unrest events labeled as PAST, ON-GOING, or FUTURE. We show that the temporal status of these events is difficult to classify because local tense and aspect cues are often lacking, time expressions are insufficient, and the linguistic contexts have rich semantic compositionality. We explore two approaches for event status classification: (1) a feature-based SVM classifier augmented with a novel induced lexicon of future-oriented verbs, such as "threatened" and "planned", and (2) a convolutional neural net. Both types of classifiers improve event status recognition over a state-of-the-art TempEval model, and our analysis offers linguistic insights into the semantic compositionality challenges for this new task.

Nested Propositions in Open Information Extraction

Nikita Bhutani, H V Jagadish, and Dragomir Radev 10:55–11:20

The challenges of Machine Reading and Knowledge Extraction at a web scale require a system capable of extracting diverse information from large, heterogeneous corpora. The Open Information Extraction (OIE) paradigm aims at extracting assertions from large corpora without requiring a vocabulary or relation-specific training data. Most systems built on this paradigm extract binary relations from arbitrary sentences, ignoring the context under which the assertions are correct and complete. They lack the expressiveness needed to properly represent and extract complex assertions commonly found in the text. To address the lack of representation power, we propose NestIE, which uses a nested representation to extract higher-order relations, and complex, interdependent assertions. Nesting the extracted propositions allows NestIE to more accurately reflect the meaning of the original sentence. Our experimental study on real-world datasets suggests that NestIE obtains comparable precision with better minimality and informativeness than existing approaches. NestIE produces 1.7-1.8 times more minimal extractions and achieves 1.1-1.2 times higher informativeness than ClausIE.

A Position Encoding Convolutional Neural Network Based on Dependency Tree for Relation Classification

Yunlun Yang, Yunhai Tong, Shulei Ma, and Zhi-Hong Deng 11:20–11:45

With the renaissance of neural network in recent years, relation classification has again become a research hotspot in natural language processing, and leveraging parse trees is a common and effective method of tackling this problem. In this work, we offer a new perspective on utilizing syntactic information of dependency parse tree and present a position encoding convolutional neural network (PECNN) based on dependency parse tree for relation classification. First, tree-based position features are proposed to encode the relative positions of words in dependency trees and help enhance the word representations. Then, based on a redefinition of "context", we design two kinds of tree-based convolution kernels for capturing the semantic and structural information provided by dependency trees. Finally, the features extracted by convolution module are fed to a classifier for labelling the semantic relations. Experiments on the benchmark dataset show that PECNN outperforms state-of-the-art approaches. We also compare the effect of different position features and visualize the influence of tree-based position feature by tracing back the convolution process.

Learning to Recognize Discontiguous Entities

Aldrian Obaja Muis and Wei Lu 11:45–12:10

This paper focuses on the study of recognizing discontiguous entities. Motivated by a previous work, we propose to use a novel hypergraph representation to jointly encode discontiguous entities of unbounded length, which can overlap with one another. To compare with existing approaches, we first formally introduce the notion of model ambiguity, which defines the difficulty level of interpreting the outputs of a model, and then formally analyze the theoretical advantages of our model over previous existing approaches based on linear-chain CRFs. Our empirical results also show that our model is able achieve significantly better results when evaluated on standard data with many discontiguous entities.

Session 1C: Psycholinguistics / Machine Learning

Room 616 AB

Modeling Human Reading with Neural Attention

Michael Hahn and Frank Keller 10:30–10:55

When humans read text, they fixate some words and skip others. However, there have been few attempts to explain skipping behavior with computational models, as most existing work has focused on predicting reading times (e.g.,~using surprisal). In this paper, we propose a novel approach that models both skipping and reading, using an unsupervised architecture that combines a neural attention with autoencoding, trained on raw text using reinforcement learning. Our model explains human reading behavior as a tradeoff between precision of language understanding (encoding the input accurately) and economy of attention (fixating as few words as possible). We evaluate the model on the Dundee eye-tracking corpus, showing that it accurately predicts skipping behavior and reading times, is competitive with surprisal, and captures known qualitative features of human reading.

Comparing Computational Cognitive Models of Generalization in a Language Acquisition Task

Libby Barak, Adele E. Goldberg, and Suzanne Stevenson 10:55–11:20

Natural language acquisition relies on appropriate generalization: the ability to produce novel sentences, while learning to restrict productions to acceptable forms in the language. Psycholinguists have proposed various properties that might play a role in guiding appropriate generalizations, looking at learning of verb alternations as a testbed. Several computational cognitive models have explored aspects of this phenomenon, but their results are hard to compare given the high variability in the linguistic properties represented in their input. In this paper, we directly compare two recent approaches, a Bayesian model and a connectionist model, in their ability to replicate human judgments of appropriate generalizations. We find that the Bayesian model more accurately mimics the judgments due to its richer learning mechanism that can exploit distributional properties of the input in a manner consistent with human behaviour.

Rationalizing Neural Predictions

Tao Lei, Regina Barzilay, and Tommi Jaakkola 11:20–11:45

Prediction without justification has limited applicability. As a remedy, we learn to extract pieces of input text as justifications – rationales – that are tailored to be short and coherent, yet sufficient for making the same prediction. Our approach combines two modular components, generator and encoder, which are trained to operate well together. The generator specifies a distribution over text fragments as candidate rationales and these are passed through the encoder for prediction. Rationales are never given during training. Instead, the model is regularized by desiderata for rationales. We evaluate the approach on multi-aspect sentiment analysis against manually annotated test cases. Our approach outperforms attention-based baseline by a significant margin. We also successfully illustrate the method on the question retrieval task.

Deep Multi-Task Learning with Shared Memory for Text Classification

Pengfei Liu, Xipeng Qiu, and Xuanjing Huang 11:45–12:10

Neural network based models have achieved impressive results on various of natural language processing (NLP) tasks. However, in previous works, most models are learned separately based on single-task supervised objectives, which often suffer from insufficient training data. In this paper, we propose two deep architectures which can be trained jointly on multiple related tasks. More specifically, we augment neural model with an external memory, which is shared by several tasks. Experiments on two groups of text classification tasks show that our proposed architectures can improve the performance of a task with the help of other related tasks.

Session 2 Overview

Track A	**Track B**	**Track C**	
Reading Comprehension and Question Answering Salon FG	*Embeddings of Linguistic Structure* Salon J	*Sentiment and Opinion Analysis* Room 616 AB	
Natural Language Comprehension with the EpiReader *Trischler, Ye, Yuan, Bachman, Sordoni, and Suleman*	Event participant modelling with neural networks *Tilk, Demberg, Sayeed, Klakow, and Thater*	Aspect Level Sentiment Classification with Deep Memory Network *Tang, Qin, and Liu*	13:40
Creating Causal Embeddings for Question Answering with Minimal Supervision *Sharp, Surdeanu, Jansen, Clark, and Hammond*	Context-Dependent Sense Embedding *Qiu, Tu, and Yu*	Lifelong-RL: Lifelong Relaxation Labeling for Separating Entities and Aspects in Opinion Targets *Shu, Liu, Xu, and Kim*	14:05
Improving Semantic Parsing via Answer Type Inference *Yavuz, Gur, Su, Srivatsa, and Yan*	Jointly Embedding Knowledge Graphs and Logical Rules *Guo, Wang, Wang, Wang, and Guo*	Learning Sentence Embeddings with Auxiliary Tasks for Cross-Domain Sentiment Classification *Yu and Jiang*	14:30
Semantic Parsing to Probabilistic Programs for Situated Question Answering *Krishnamurthy, Tafjord, and Kembhavi*	Learning Connective-based Word Representations for Implicit Discourse Relation Identification *Braud and Denis*	Attention-based LSTM Network for Cross-Lingual Sentiment Classification *Zhou, Wan, and Xiao*	14:55

Session 2A: Reading Comprehension and Question Answering

Salon FG

Natural Language Comprehension with the EpiReader

Adam Trischler, Zheng Ye, Xingdi Yuan, Philip Bachman, Alessandro Sordoni, and Kaheer Suleman 13:40–14:05

We present EpiReader, a novel model for machine comprehension of text. Machine comprehension of unstructured, real-world text is a major research goal for natural language processing. Current tests of machine comprehension pose questions whose answers can be inferred from some supporting text, and evaluate a model's response to the questions. EpiReader is an end-to-end neural model comprising two components: the first component proposes a small set of candidate answers after comparing a question to its supporting text, and the second component formulates hypotheses using the proposed candidates and the question, then reranks the hypotheses based on their estimated concordance with the supporting text. We present experiments demonstrating that EpiReader sets a new state-of-the-art on the CNN and Children's Book Test benchmarks, outperforming previous neural models by a significant margin.

Creating Causal Embeddings for Question Answering with Minimal Supervision

Rebecca Sharp, Mihai Surdeanu, Peter Jansen, Peter Clark, and Michael Hammond 14:05–14:30

A common model for question answering (QA) is that a good answer is one that is closely related to the question, where relatedness is often determined using general-purpose lexical models such as word embeddings. We argue that a better approach is to look for answers that are related to the question in a relevant way, according to the information need of the question, which may be determined through task-specific embeddings. With causality as a use case, we implement this insight in three steps. First, we generate causal embeddings cost-effectively by bootstrapping cause-effect pairs extracted from free text using a small set of seed patterns. Second, we train dedicated embeddings over this data, by using task-specific contexts, i.e., the context of a cause is its effect. Finally, we extend a state-of-the-art reranking approach for QA to incorporate these causal embeddings. We evaluate the causal embedding models both directly with a casual implication task, and indirectly, in a downstream causal QA task using data from Yahoo! Answers. We show that explicitly modeling causality improves performance in both tasks. In the QA task our best model achieves 37.3% P1, significantly outperforming a strong baseline by 7.7% (relative).

Improving Semantic Parsing via Answer Type Inference

Semih Yavuz, Izzeddin Gur, Yu Su, Mudhakar Srivatsa, and Xifeng Yan 14:30–14:55

In this work, we show the possibility of inferring the answer type before solving a factoid question and leveraging the type information to improve semantic parsing. By replacing the topic entity in a question with its type, we are able to generate an abstract form of the question, whose answer corresponds to the answer type of the original question. A bidirectional LSTM model is built to train over the abstract form of questions and infer their answer types. It is also observed that if we convert a question into a statement form, our LSTM model achieves better accuracy. Using the predicted type information to rerank the logical forms returned by AgendaIL, one of the leading semantic parsers, we are able to improve the F1-score from 49.7 to 52.6 on the WEBQUESTIONS data.

Semantic Parsing to Probabilistic Programs for Situated Question Answering

Jayant Krishnamurthy, Oyvind Tafjord, and Aniruddha Kembhavi 14:55–15:20

Situated question answering is the problem of answering questions about an environment such as an image or diagram. This problem requires jointly interpreting a question and an environment using background knowledge to select the correct answer. We present Parsing to Probabilistic Programs (P3), a novel situated question answering model that can use background knowledge and global features of the question/environment interpretation while retaining efficient approximate inference. Our key insight is to treat semantic parses as probabilistic programs that execute nondeterministically and whose possible executions represent environmental uncertainty. We evaluate our approach on a new, publicly-released data set of 5000 science diagram questions, outperforming several competitive classical and neural baselines.

Session 2B: Embeddings of Linguistic Structure

Salon J

Event participant modelling with neural networks

Ottokar Tilk, Vera Demberg, Asad Sayeed, Dietrich Klakow, and Stefan Thater 13:40–14:05

A common problem in cognitive modelling is lack of access to accurate broad-coverage models of event-level surprisal. As shown in, e.g., Bicknell et al. (2010), event-level knowledge does affect human expectations for verbal arguments. For example, the model should be able to predict that "mechanics" are likely to check "tires", while "journalists" are more likely to check "typos". Similarly, we would like to predict what locations are likely for "playing football" or "playing flute" in order to estimate the surprisal of actually-encountered locations. Furthermore, such a model can be used to provide a probability distribution over fillers for a thematic role which is not mentioned in the text at all. To this end, we train two neural network models (an incremental one and a non-incremental one) on large amounts of automatically role-labelled text. Our models are probabilistic and can handle several roles at once, which also enables them to learn interactions between different role fillers. Evaluation shows a drastic improvement over current state-of-the-art systems on modelling human thematic fit judgements, and we demonstrate via a sentence similarity task that the system learns highly useful embeddings.

Context-Dependent Sense Embedding

Lin Qiu, Kewei Tu, and Yong Yu 14:05–14:30

Word embedding has been widely studied and proven helpful in solving many natural language processing tasks. However, the ambiguity of natural language is always a problem on learning high quality word embeddings. A possible solution is sense embedding which trains embedding for each sense of words instead of each word. Some recent work on sense embedding uses context clustering methods to determine the senses of words, which is heuristic in nature. Other work creates a probabilistic model and performs word sense disambiguation and sense embedding iteratively. However, most of the previous work has the problems of learning sense embeddings based on imperfect word embeddings as well as ignoring the dependency between sense choices of neighboring words. In this paper, we propose a novel probabilistic model for sense embedding that is not based on problematic word embedding of polysemous words and takes into account the dependency between sense choices. Based on our model, we derive a dynamic programming inference algorithm and an Expectation-Maximization style learning algorithm. The empirical studies show that our model outperforms the state-of-the-art model on a word sense induction task by a 13% relative gain.

Jointly Embedding Knowledge Graphs and Logical Rules

Shu Guo, Quan Wang, Lihong Wang, Bin Wang, and Li Guo 14:30–14:55

Embedding knowledge graphs into continuous vector spaces has recently attracted increasing interest. Most existing methods perform the embedding task using only fact triples. Logical rules, although containing rich background information, have not been well studied in this task. This paper proposes a novel method of jointly embedding knowledge graphs and logical rules. The key idea is to represent and model triples and rules in a unified framework. Specifically, triples are represented as atomic formulae and modeled by the translation assumption, while rules represented as complex formulae and modeled by t-norm fuzzy logics. Embedding then amounts to minimizing a global loss over both atomic and complex formulae. In this manner, we learn embeddings compatible not only with triples but also with rules, which will certainly be more predictive for knowledge acquisition and inference. We evaluate our method with link prediction and triple classification tasks. Experimental results show that joint embedding brings significant and consistent improvements over state-of-the-art methods. Particularly, it enhances the prediction of new facts which cannot even be directly inferred by pure logical inference, demonstrating the capability of our method to learn more predictive embeddings.

Learning Connective-based Word Representations for Implicit Discourse Relation Identification

Chloé Braud and Pascal Denis 14:55–15:20

We introduce a simple semi-supervised approach to improve implicit discourse relation identification. This approach harnesses large amounts of automatically extracted discourse connectives along with their arguments to construct new distributional word representations. Specifically, we represent words in the space of discourse connectives as a way to directly encode their rhetorical function. Experiments on the Penn Discourse Treebank demonstrate the effectiveness of these task-tailored representations in predicting implicit discourse relations. Our results indeed show that, despite their simplicity, these connective-based representations outperform various off-the-shelf word embeddings, and achieve state-of-the-art performance on this problem.

Session 2C: Sentiment and Opinion Analysis

Room 616 AB

Aspect Level Sentiment Classification with Deep Memory Network

Duyu Tang, Bing Qin, and Ting Liu 13:40–14:05

We introduce a deep memory network for aspect level sentiment classification. Unlike feature-based SVM and sequential neural models such as LSTM, this approach explicitly captures the importance of each context word when inferring the sentiment polarity of an aspect. Such importance degree and text representation are calculated with multiple computational layers, each of which is a neural attention model over an external memory. Experiments on laptop and restaurant datasets demonstrate that our approach performs comparable to state-of-art feature based SVM system, and substantially better than LSTM and attention-based LSTM architectures. On both datasets we show that multiple computational layers could improve the performance. Moreover, our approach is also fast. The deep memory network with 9 layers is 15 times faster than LSTM with a CPU implementation.

Lifelong-RL: Lifelong Relaxation Labeling for Separating Entities and Aspects in Opinion Targets

Lei Shu, Bing Liu, Hu Xu, and Annice Kim 14:05–14:30

It is well-known that opinions have targets. Extracting such targets is an important problem of opinion mining because without knowing the target of an opinion, the opinion is of limited use. So far many algorithms have been proposed to extract opinion targets. However, an opinion target can be an entity or an aspect (part or attribute) of an entity. An opinion about an entity is an opinion about the entity as a whole, while an opinion about an aspect is just an opinion about that specific attribute or aspect of an entity. Thus, opinion targets should be separated into entities and aspects before use because they represent very different things about opinions. This paper proposes a novel algorithm, called Lifelong-RL, to solve the problem based on lifelong machine learning and relaxation labeling. Extensive experiments show that the proposed algorithm Lifelong-RL outperforms baseline methods markedly.

Learning Sentence Embeddings with Auxiliary Tasks for Cross-Domain Sentiment Classification

Jianfei Yu and Jing Jiang 14:30–14:55

In this paper, we study cross-domain sentiment classification with neural network architectures. We borrow the idea from Structural Correspondence Learning and use two auxiliary tasks to help induce a sentence embedding that supposedly works well across domains for sentiment classification. We also propose to jointly learn this sentence embedding together with the sentiment classifier itself. Experiment results demonstrate that our proposed joint model outperforms several state-of-the-art methods on five benchmark datasets.

Attention-based LSTM Network for Cross-Lingual Sentiment Classification

Xinjie Zhou, Xiaojun Wan, and Jianguo Xiao 14:55–15:20

Most of the state-of-the-art sentiment classification methods are based on supervised learning algorithms which require large amounts of manually labeled data. However, the labeled resources are usually imbalanced in different languages. Cross-lingual sentiment classification tackles the problem by adapting the sentiment resources in a resource-rich language to resource-poor languages. In this study, we propose an attention-based bilingual representation learning model which learns the distributed semantics of the documents in both the source and the target languages. In each language, we use Long Short Term Memory (LSTM) network to model the documents, which has been proved to be very effective for word sequences. Meanwhile, we propose a hierarchical attention mechanism for the bilingual LSTM network. The sentence-level attention model learns which sentences of a document are more important for determining the overall sentiment while the word-level attention model learns which words in each sentence are decisive. The proposed model achieves good results on a benchmark dataset using English as the source language and Chinese as the target language.

Session 3 Overview

Track A	**Track B**	**Track C**	
Neural Machine Translation Salon FG	*Semi-supervised and Minimally Supervised Learning* Salon J	*Summarization and Generation* Room 616 AB	
Deep Recurrent Models with Fast-Forward Connections for Neural Machine Translation *Zhou, Cao, Wang, Li, and Xu*	Semi-Supervised Learning of Sequence Models with Method of Moments *Marinho, Martins, Cohen, and Smith*	Language as a Latent Variable: Discrete Generative Models for Sentence Compression *Miao and Blunsom*	15:50
Neural versus Phrase-Based Machine Translation Quality: a Case Study *Bentivogli, Bisazza, Cettolo, and Federico*	Minimally supervised models for number normalization *Gorman and Sproat*	Globally Coherent Text Generation with Neural Checklist Models *Kiddon, Zettlemoyer, and Choi*	16:15
Zero-Resource Translation with Multi-Lingual Neural Machine Translation *Firat, Sankaran, Al-Onaizan, Yarman Vural, and Cho*	Learning from Explicit and Implicit Supervision Jointly For Algebra Word Problems *Upadhyay, Chang, Chang, and Yih*	A Dataset and Evaluation Metrics for Abstractive Compression of Sentences and Short Paragraphs *Toutanova, Brockett, Tran, and Amershi*	16:40
Memory-enhanced Decoder for Neural Machine Translation *Wang, Lu, Li, and Liu*	TweeTime : A Minimally Supervised Method for Recognizing and Normalizing Time Expressions in Twitter *Tabassum, Ritter, and Xu*	PaCCSS-IT: A Parallel Corpus of Complex-Simple Sentences for Automatic Text Simplification *Brunato, Cimino, Dell'Orletta, and Venturi*	17:05

Session 3A: Neural Machine Translation

Salon FG

Deep Recurrent Models with Fast-Forward Connections for Neural Machine Translation

Jie Zhou, Ying Cao, Xuguang Wang, Peng Li, and Wei Xu 15:50–16:15

Neural machine translation (NMT) aims at solving machine translation (MT) problems using neural networks and has exhibited promising results in recent years. However, most of the existing NMT models are shallow and there is still a performance gap between a single NMT model and the best conventional MT system. In this work, we introduce a new type of linear connections, named fast-forward connections, based on deep Long Short-Term Memory (LSTM) networks, and an interleaved bi-directional architecture for stacking the LSTM layers. Fast-forward connections play an essential role in propagating the gradients and building a deep topology of depth 16. On the WMT'14 English-to-French task, we achieve BLEU=37.7 with a single attention model, which outperforms the corresponding single shallow model by 6.2 BLEU points. This is the first time that a single NMT model achieves state-of-the-art performance and outperforms the best conventional model by 0.7 BLEU points. We can still achieve BLEU=36.3 even without using an attention mechanism. After special handling of unknown words and model ensembling, we obtain the best score reported to date on this task with BLEU=40.4. Our models are also validated on the more difficult WMT'14 English-to-German task.

Neural versus Phrase-Based Machine Translation Quality: a Case Study

Luisa Bentivogli, Arianna Bisazza, Mauro Cettolo, and Marcello Federico 16:15–16:40

Within the field of Statistical Machine Translation (SMT), the neural approach (NMT) has recently emerged as the first technology able to challenge the long-standing dominance of phrase-based approaches (PBMT). In particular, at the IWSLT 2015 evaluation campaign, NMT outperformed well established state-of-the-art PBMT systems on English-German, a language pair known to be particularly hard because of morphology and syntactic differences. To understand in what respects NMT provides better translation quality than PBMT, we perform a detailed analysis of neural vs. phrase-based SMT outputs, leveraging high quality post-edits performed by professional translators on the IWSLT data. For the first time, our analysis provides useful insights on what linguistic phenomena are best modeled by neural models – such as the reordering of verbs – while pointing out other aspects that remain to be improved.

Zero-Resource Translation with Multi-Lingual Neural Machine Translation

Orhan Firat, Baskaran Sankaran, Yaser Al-Onaizan, Fatos T. Yarman Vural, and Kyunghyun Cho 16:40–17:05

In this paper, we propose a novel finetuning algorithm for the recently introduced multi-way, multilingual neural machine translate that enables zero-resource machine translation. When used together with novel many-to-one translation strategies, we empirically show that this finetuning algorithm allows the multi-way, multilingual model to translate a zero-resource language pair (1) as well as a single-pair neural translation model trained with up to 1M direct parallel sentences of the same language pair and (2) better than pivot-based translation strategy, while keeping only one additional copy of attention-related parameters.

Memory-enhanced Decoder for Neural Machine Translation

Mingxuan Wang, Zhengdong Lu, Hang Li, and Qun Liu 17:05–17:30

We propose to enhance the RNN decoder in a neural machine translator (NMT) with external memory, as a natural but powerful extension to the state in the decoding RNN. This memory-enhanced RNN decoder is called MemDec. At each time during decoding, MemDec will read from this memory and write to this memory once, both with content-based addressing. Unlike the unbounded memory in previous work[**RNNsearch**] to store the representation of source sentence, the memory in MemDec is a matrix with pre-determined size designed to better capture the information important for the decoding process at each time step. Our empirical study on Chinese-English translation shows that it can improve by 4.8 BLEU upon Groundhog and 5.3 BLEU upon on Moses, yielding the best performance achieved with the same training set.

Session 3B: Semi-supervised and Minimally Supervised Learning

Salon J

Semi-Supervised Learning of Sequence Models with Method of Moments

Zita Marinho, André F. T. Martins, Shay B. Cohen, and Noah A. Smith 15:50–16:15

We propose a fast and scalable method for semi-supervised learning of sequence models, based on anchor words and moment matching. Our method can handle hidden Markov models with feature-based log-linear emissions. Unlike other semi-supervised methods, no decoding passes are necessary on the unlabeled data and no graph needs to be constructed—only one pass is necessary to collect moment statistics. The model parameters are estimated by solving a small quadratic program for each feature. Experiments on part-of-speech (POS) tagging for Twitter and for a low resource language (Malagasy) show that our method can learn from very few annotated sentences.

Minimally supervised models for number normalization

Kyle Gorman and Richard Sproat 16:15–16:40

We propose two models for verbalizing numbers, a key component in speech recognition and synthesis systems. The first model uses an end-to-end recurrent neural network. The second model, drawing inspiration from the linguistics literature, uses finite-state transducers constructed with a minimal amount of training data. While both models achieve near-perfect performance, the latter model can be trained using several orders of magnitude less data than the former, making it particularly useful for low-resource languages.

Learning from Explicit and Implicit Supervision Jointly For Algebra Word Problems

Shyam Upadhyay, Ming-Wei Chang, Kai-Wei Chang, and Wen-tau Yih 16:40–17:05

Automatically solving algebra word problems has raised considerable interest recently. Existing state-of-the-art approaches mainly rely on learning from human annotated equations. In this paper, we demonstrate that it is possible to efficiently mine algebra problems and their numerical solutions with little to no manual effort. To leverage the mined dataset, we propose a novel structured-output learning algorithm that aims to learn from both explicit (e.g., equations) and implicit (e.g., solutions) supervision signals jointly. Enabled by this new algorithm, our model gains 4.6% absolute improvement in accuracy on the ALG-514 benchmark compared to the one without using implicit supervision. The final model also outperforms the current state-of-the-art approach by 3%.

TweeTime : A Minimally Supervised Method for Recognizing and Normalizing Time Expressions in Twitter

Jeniya Tabassum, Alan Ritter, and Wei Xu 17:05–17:30

We describe TweeTIME, a temporal tagger for recognizing and normalizing time expressions in Twitter. Most previous work in social media analysis has to rely on temporal resolvers that are designed for well-edited text, and therefore suffer from the reduced performance due to domain mismatch. We present a minimally supervised method that learns from large quantities of unlabeled data and requires no hand-engineered rules or hand-annotated training corpora. TweeTIME achieves 0.68 F1 score on the end-to-end task of resolving date expressions, outperforming a broad range of state-of-the-art systems.

Session 3C: Summarization and Generation

Room 616 AB

Language as a Latent Variable: Discrete Generative Models for Sentence Compression

Yishu Miao and Phil Blunsom 15:50–16:15

In this work we explore deep generative models of text in which the latent representation of a document is itself drawn from a discrete language model distribution. We formulate a variational auto-encoder for inference in this model and apply it to the task of compressing sentences. In this application the generative model first draws a latent summary sentence from a background language model, and then subsequently draws the observed sentence conditioned on this latent summary. In our empirical evaluation we show that generative formulations of both abstractive and extractive compression yield state-of-the-art results when trained on a large amount of supervised data. Further, we explore semi-supervised compression scenarios where we show that it is possible to achieve performance competitive with previously proposed supervised models while training on a fraction of the supervised data.

Globally Coherent Text Generation with Neural Checklist Models

Chloé Kiddon, Luke Zettlemoyer, and Yejin Choi 16:15–16:40

Recurrent neural networks can generate locally coherent text but often have difficulties representing what has already been generated and what still needs to be said – especially when constructing long texts. We present the neural checklist model, a recurrent neural network that models global coherence by storing and updating an agenda of text strings which should be mentioned somewhere in the output. The model generates output by dynamically adjusting the interpolation among a language model and a pair of attention models that encourage references to agenda items. Evaluations on cooking recipes and dialogue system responses demonstrate high coherence with greatly improved semantic coverage of the agenda.

A Dataset and Evaluation Metrics for Abstractive Compression of Sentences and Short Paragraphs

Kristina Toutanova, Chris Brockett, Ke M. Tran, and Saleema Amershi 16:40–17:05

We introduce a manually-created, multi-reference dataset for abstractive sentence and short paragraph compression. First, we examine the impact of single- and multi-sentence level editing operations on human compression quality as found in this corpus. We observe that substitution and rephrasing operations are more meaning preserving than other operations, and that compressing in context improves quality. Second, we systematically explore the correlations between automatic evaluation metrics and human judgments of meaning preservation and grammaticality in the compression task, and analyze the impact of the linguistic units used and precision versus recall measures on the quality of the metrics. Multi-reference evaluation metrics are shown to offer significant advantage over single reference-based metrics.

PaCCSS-IT: A Parallel Corpus of Complex-Simple Sentences for Automatic Text Simplification

Dominique Brunato, Andrea Cimino, Felice Dell'Orletta, and Giulia Venturi 17:05–17:30

In this paper we present PaCCSS–IT, a Parallel Corpus of Complex–Simple Sentences for ITalian. To build the resource we develop a new method for automatically acquiring a corpus of complex–simple paired sentences able to intercept structural transformations and particularly suitable for text simplification. The method requires a wide amount of texts that can be easily extracted from the web making it suitable also for less–resourced languages. We test it on the Italian language making available the biggest Italian corpus for automatic text simplification.

Poster Spotlight A (Half-minute Madness)

Chair: Brendan O'Connor Joel Tetreault

Salon H and J *17:45–18:15*

Poster Session A

Salon H and J *18:15–20:15*

[Discourse & Dialogue] [ID:L01]

Discourse Parsing with Attention-based Hierarchical Neural Networks

Qi Li, Tianshi Li, and Baobao Chang

RST-style document-level discourse parsing remains a difficult task and efficient deep learning models on this task have rarely been presented. In this paper, we propose an attention-based hierarchical neural network model for discourse parsing. We also incorporate tensor-based transformation function to model complicated feature interactions. Experimental results show that our approach obtains comparable performance to the contemporary state-of-the-art systems with little manual feature engineering.

[Discourse & Dialogue] [ID:L02]

Multi-view Response Selection for Human-Computer Conversation

Xiangyang Zhou, Daxiang Dong, Hua Wu, Shiqi Zhao, Dianhai Yu, Hao Tian, Xuan Liu, and Rui Yan

In this paper, we study the task of response selection for multi-turn human-computer conversation. Previous approaches take word as a unit and view context and response as sequences of words. This kind of approaches do not explicitly take each utterance as a unit, therefore it is difficult to catch utterance-level discourse information and dependencies. In this paper, we propose a multi-view response selection model that integrates information from two different views, i.e., word sequence view and utterance sequence view. We jointly model the two views via deep neural networks. Experimental results on a public corpus for context-sensitive response selection demonstrate the effectiveness of the proposed multi-view model, which significantly outperforms other single-view baselines.

[Discourse & Dialogue] [ID:L03]

Variational Neural Discourse Relation Recognizer

Biao Zhang, Deyi Xiong, Qun Liu, Rongrong Ji, Hong Duan, and Min Zhang

Implicit discourse relation recognition is a crucial component for automatic discourselevel analysis and nature language understanding. Previous studies exploit discriminative models that are built on either powerful manual features or deep discourse representations. In this paper, instead, we explore generative models and propose a variational neural discourse relation recognizer. We refer to this model as VarNDRR. VarNDRR establishes a directed probabilistic model with a latent continuous variable that generates both a discourse and the relation between the two arguments of the discourse. In order to perform efficient inference and learning, we introduce neural discourse relation models to approximate the prior and posterior distributions of the latent variable, and employ these approximated distributions to optimize a reparameterized variational lower bound. This allows VarNDRR to be trained with standard stochastic gradient methods. Experiments on the benchmark data set show that VarNDRR can achieve comparable results against stateof- the-art baselines without using any manual features.

[Information Extraction] [ID:L04]

Event Detection and Co-reference with Minimal Supervision

Haoruo Peng, Yangqiu Song, and Dan Roth

An important aspect of natural language understanding involves recognizing and categorizing events and the relations among them. However, these tasks are quite subtle and annotating training data for machine learning based approaches is an expensive task, resulting in supervised systems that attempt to learn complex models from small amounts of data, which they over-fit. This paper addresses this challenge by developing an event detection and co-reference system with minimal supervision, in the form of a few event examples. We view these tasks as semantic similarity problems between event mentions or event mentions and an ontology of types, thus facilitating the use of large amounts of out of domain text data. Notably, our semantic relatedness function exploits the structure of the text by making use of a semantic-role-labeling based representation of an event. We show that our approach to event detection is competitive with the top supervised methods. More significantly, we outperform state-of-the-art supervised methods for event co-reference on benchmark data sets, and support significantly better transfer across domains.

[Information Extraction] [ID:L05]

Learning Term Embeddings for Taxonomic Relation Identification Using Dynamic Weighting Neural Network

Tuan Luu Anh, Yi Tay, Siu Cheung Hui, and See Kiong Ng

Taxonomic relation identification aims to recognize the 'is-a' relation between two terms. Previous works on identifying taxonomic relations are mostly based on statistical and linguistic approaches, but the accuracy of these approaches is far from satisfactory. In this paper, we propose a novel supervised learning approach for identifying taxonomic relations using term embeddings. For this purpose, we first design a dynamic weighting neural network to learn term embeddings based on not only the hypernym and hyponym terms, but also the contextual information between them. We then apply such embeddings as features to identify taxonomic relations using a supervised method. The experimental results show that our proposed approach significantly outperforms other state-of-the-art methods by 9% to 13% in terms of accuracy for both general and specific domain datasets.

[Information Extraction] [ID:L06]

Relation Schema Induction using Tensor Factorization with Side Information

Madhav Nimishakavi, Uday Singh Saini, and Partha Talukdar

Given a set of documents from a specific domain (e.g., medical research journals), how do we automatically build a Knowledge Graph (KG) for that domain? Automatic identification of relations and their schemas, i.e., type signature of arguments of relations (e.g., undergo(Patient, Surgery)), is an important first step towards this goal. We refer to this problem as Relation Schema Induction (RSI). In this paper, we propose Schema Induction using Coupled Tensor Factorization(SICTF), a novel tensor factorization method for relation schema induction. SICTF factorizes Open Information Extraction (OpenIE) triples extracted from a domain corpus along with additional side information in a principled way to induce relation schemas. To the best of our knowledge, this is the first application of tensor factorization for the RSI problem. Through extensive experiments on multiple real-world datasets, we find that SICTF is not only more accurate than state-of-the-art baselines, but also significantly faster (about 14x faster).

[Information Extraction] [ID:L07]

Supervised Distributional Hypernym Discovery via Domain Adaptation

Luis Espinosa Anke, Jose Camacho-Collados, Claudio Delli Bovi, and Horacio Saggion

Lexical taxonomies are graph-like hierarchical structures that provide a formal representation of knowledge. Most knowledge graphs to date rely on is-a (hypernymic) relations as the backbone of their semantic structure. In this paper, we propose a supervised distributional framework for hypernym discovery which operates at the sense level, enabling large-scale automatic acquisition of disambiguated taxonomies. By exploiting semantic regularities between hyponyms and hypernyms in embeddings spaces, and integrating a domain clustering algorithm, our model becomes sensitive to the target data. We evaluate several configurations of our approach, training with information derived from a manually created knowledge base, along with hypernymic relations obtained from Open Information Extraction systems. The integration of both sources of knowledge yields the best overall results according to both automatic and manual evaluation on ten different domains.

[Language Modeling] [ID:L08]

Latent Tree Language Model

Tomáš Brychcín

In this paper we introduce Latent Tree Language Model (LTLM), a novel approach to language modeling that encodes syntax and semantics of a given sentence as a tree of word roles. The learning phase iteratively updates the trees by moving nodes according to Gibbs sampling. We introduce two algorithms to infer a tree for a given sentence. The first one is based on Gibbs sampling. It is fast, but does not guarantee to find the most probable tree. The second one is based on dynamic programming. It is slower,

but guarantees to find the most probable tree. We provide comparison of both algorithms. We combine LTLM with 4-gram Modified Kneser-Ney language model via linear interpolation. Our experiments with English and Czech corpora show significant perplexity reductions (up to 46% for English and 49% for Czech) compared with standalone 4-gram Modified Kneser-Ney language model.

[Language and Vision] [ID:L09]

Comparing Data Sources and Architectures for Deep Visual Representation Learning in Semantics

Douwe Kiela, Anita Lilla Verő, and Stephen Clark

Multi-modal distributional models learn grounded representations for improved performance in semantics tasks. Deep visual representations, learned using convolutional neural networks, have been shown to achieve particularly high performance. In this study, we systematically compare deep visual representation learning techniques, experimenting with three well-known network architectures. In addition, we explore the various data sources that can be used for retrieving relevant images, showing that images from search engines perform as well as, or better than, manually crafted resources such as ImageNet. Furthermore, we explore the optimal number of images and the multi-lingual applicability of multi-modal semantics. We hope that these findings can serve as a guide for future research in the field.

[Language and Vision] [ID:L10]

Multimodal Compact Bilinear Pooling for Visual Question Answering and Visual Grounding

Akira Fukui, Dong Huk Park, Daylen Yang, Anna Rohrbach, Trevor Darrell, and Marcus Rohrbach

Modeling textual or visual information with vector representations trained from large language or visual datasets has been successfully explored in recent years. However, tasks such as visual question answering require combining these vector representations with each other. Approaches to multimodal pooling include element-wise product or sum, as well as concatenation of the visual and textual representations. We hypothesize that these methods are not as expressive as an outer product of the visual and textual vectors. As the outer product is typically infeasible due to its high dimensionality, we instead propose utilizing Multimodal Compact Bilinear pooling (MCB) to efficiently and expressively combine multimodal features. We extensively evaluate MCB on the visual question answering and grounding tasks. We consistently show the benefit of MCB over ablations without MCB. For visual question answering, we present an architecture which uses MCB twice, once for predicting attention over spatial features and again to combine the attended representation with the question representation. This model outperforms the state-of-the-art on the Visual7W dataset and the VQA challenge.

[Machine Learning] [ID:L11]

The Structured Weighted Violations Perceptron Algorithm

Rotem Dror and Roi Reichart

We present the Structured Weighted Violations Perceptron (SWVP) algorithm, a new structured prediction algorithm that generalizes the Collins Structured Perceptron (CSP, (Collins, 2002)). Unlike CSP, the update rule of SWVP explicitly exploits the internal structure of the predicted labels. We prove the convergence of SWVP for linearly separable training sets, provide mistake and generalization bounds, and show that in the general case these bounds are tighter than those of the CSP special case. In synthetic data experiments with data drawn from an HMM, various variants of SWVP substantially outperform its CSP special case. SWVP also provides encouraging initial dependency parsing results.

[Machine Learning] [ID:L12]

How Transferable are Neural Networks in NLP Applications?

Lili Mou, Zhao Meng, Rui Yan, Ge Li, Yan Xu, Lu Zhang, and Zhi Jin

Transfer learning is aimed to make use of valuable knowledge in a source domain to help model performance in a target domain. It is particularly important to neural networks, which are very likely to be overfitting. In some fields like image processing, many studies have shown the effectiveness of neural network-based transfer learning. For neural NLP, however, existing studies have only casually applied transfer learning, and conclusions are inconsistent. In this paper, we conduct systematic case studies and provide an illuminating picture on the transferability of neural networks in NLP.

[Machine Learning] [ID:L13]

Morphological Priors for Probabilistic Neural Word Embeddings

Parminder Bhatia, Robert Guthrie, and Jacob Eisenstein

Word embeddings allow natural language processing systems to share statistical information across related words. These embeddings are typically based on distributional statistics, making it difficult for

them to generalize to rare or unseen words. We propose to improve word embeddings by incorporating morphological information, capturing shared sub-word features. Unlike previous work that constructs word embeddings directly from morphemes, we combine morphological and distributional information in a unified probabilistic framework, in which the word embedding is a latent variable. The morphological information provides a prior distribution on the latent word embeddings, which in turn condition a likelihood function over an observed corpus. This approach yields improvements on intrinsic word similarity evaluations, and also in the downstream task of part-of-speech tagging.

[Machine Translation] [ID:L14]

Automatic Cross-Lingual Similarization of Dependency Grammars for Tree-based Machine Translation

Wenbin Jiang, Wen Zhang, Jinan Xu, and Rangjia Cai

Structural isomorphism between languages benefits the performance of cross-lingual applications. We propose an automatic algorithm for cross-lingual similarization of dependency grammars, which automatically learns grammars with high cross-lingual similarity. The algorithm similarizes the annotation styles of the dependency grammars for two languages in the level of classification decisions, and gradually improves the cross-lingual similarity without losing linguistic knowledge resorting to iterative cross-lingual cooperative learning. The dependency grammars given by cross-lingual similarization have much higher cross-lingual similarity while maintaining non-triviality. As applications, the cross-lingually similarized grammars significantly improve the performance of dependency tree-based machine translation.

[Machine Translation] [ID:L15]

IRT-based Aggregation Model of Crowdsourced Pairwise Comparison for Evaluating Machine Translations

Naoki Otani, Toshiaki Nakazawa, Daisuke Kawahara, and Sadao Kurohashi

Recent work on machine translation has used crowdsourcing to reduce costs of manual evaluations. However, crowdsourced judgments are often biased and inaccurate. In this paper, we present a statistical model that aggregates many manual pairwise comparisons to robustly measure a machine translation system's performance. Our method applies graded response model from item response theory (IRT), which was originally developed for academic tests. We conducted experiments on a public dataset from the Workshop on Statistical Machine Translation 2013, and found that our approach resulted in highly interpretable estimates and was less affected by noisy judges than previously proposed methods.

[Machine Translation] [ID:L16]

Variational Neural Machine Translation

Biao Zhang, Deyi Xiong, Hong Duan, and Min Zhang

Models of neural machine translation are often from a discriminative family of encoderdecoders that learn a conditional distribution of a target sentence given a source sentence. In this paper, we propose a variational model to learn this conditional distribution for neural machine translation: a variational encoderdecoder model that can be trained end-to-end. Different from the vanilla encoder-decoder model that generates target translations from hidden representations of source sentences alone, the variational model introduces a continuous latent variable to explicitly model underlying semantics of source sentences and to guide the generation of target translations. In order to perform efficient posterior inference and large-scale training, we build a neural posterior approximator conditioned on both the source and the target sides, and equip it with a reparameterization technique to estimate the variational lower bound. Experiments on both Chinese-English and English- German translation tasks show that the proposed variational neural machine translation achieves significant improvements over the vanilla neural machine translation baselines.

[Machine Translation] [ID:L17]

Towards a Convex HMM Surrogate for Word Alignment

Andrei Simion, Michael Collins, and Cliff Stein

Among the alignment models used in statistical machine translation (SMT), the hidden Markov model (HMM) is arguably the most elegant: it performs consistently better than IBM Model 3 and is very close in performance to the much more complex IBM Model 4. In this paper we discuss a model which combines the structure of the HMM and IBM Model 2. Using this surrogate, our experiments show that we can attain a similar level of alignment quality as the HMM model implemented in GIZA++ [**och03**]. For this model, we derive its convex relaxation and show that it too has strong performance despite not having the local optima problems of non-convex objectives. In particular, the word alignment quality of this new convex model is significantly above that of the standard IBM Models 2 and 3, as well as the popular (and still non-convex) IBM Model 2 variant of [**dyer13**].

[Question Answering] [ID:L18]

Solving Verbal Questions in IQ Test by Knowledge-Powered Word Embedding
Huazheng Wang, Fei Tian, Bin Gao, Chengjieren Zhu, Jiang Bian, and Tie-Yan Liu

Verbal comprehension questions appear very frequently in Intelligence Quotient (IQ) tests, which measure human's verbal ability including the understanding of the words with multiple senses, the synonyms and antonyms, and the analogies among words. In this work, we explore whether such tests can be solved automatically by the deep learning technologies for text data. We found that the task was quite challenging, and simply applying existing technologies like word embedding could not achieve a good performance, due to the multiple senses of words and the complex relations among words. To tackle these challenges, we propose a novel framework to automatically solve the verbal IQ questions by leveraging improved word embedding by jointly considering the multi-sense nature of words and the relational information among words. Experimental results have shown that the proposed framework can not only outperform existing methods for solving verbal comprehension questions but also exceed the average performance of the Amazon Mechanical Turk workers involved in the study.

[Question Answering] [ID:L19]

Long Short-Term Memory-Networks for Machine Reading
Jianpeng Cheng, Li Dong, and Mirella Lapata

In this paper we address the question of how to render sequence-level networks better at handling structured input. We propose a machine reading simulator which processes text incrementally from left to right and performs shallow reasoning with memory and attention. The reader extends the Long Short-Term Memory architecture with a memory network in place of a single memory cell. This enables adaptive memory usage during recurrence with neural attention, offering a way to weakly induce relations among tokens. The system is initially designed to process a single sequence but we also demonstrate how to integrate it with an encoder-decoder architecture. Experiments on language modeling, sentiment analysis, and natural language inference show that our model matches or outperforms the state of the art.

[Question Answering] [ID:L20]

On Generating Characteristic-rich Question Sets for QA Evaluation
Yu Su, Huan Sun, Brian Sadler, Mudhakar Srivatsa, Izzeddin Gur, Zenghui Yan, and Xifeng Yan

We present a semi-automated framework for constructing factoid question answering (QA) datasets, where an array of question characteristics are formalized, including structure complexity, function, commonness, answer cardinality, and paraphrasing. Instead of collecting questions and manually characterizing them, we employ a reverse procedure, first generating a kind of graph-structured logical forms from a knowledge base, and then converting them into questions. Our work is the first to generate questions with explicitly specified characteristics for QA evaluation. We construct a new QA dataset with over 5,000 logical form-question pairs, associated with answers from the knowledge base, and show that datasets constructed in this way enable fine-grained analyses of QA systems. The dataset can be found in https://github.com/ysu1989/GraphQuestions.

[Question Answering] [ID:L21]

Learning to Translate for Multilingual Question Answering
Ferhan Ture and Elizabeth Boschee

In multilingual question answering, either the question needs to be translated into the document language, or vice versa. In addition to direction, there are multiple ways to perform the translation, four of which we explore in this paper: word-based, 10-best, context-based, and grammar-based. We build a feature for each combination of translation direction and method, and train a model that learns optimal feature weights. On a large forum dataset consisting of posts in English, Arabic, and Chinese, our novel learn-to-translate approach was more effective than a typical approach ($p<0.05$): translating all text into English, then training a classifier based only on English (original or translated) text.

[Question Answering] [ID:L22]

A Semiparametric Model for Bayesian Reader Identification
Ahmed Abdelwahab, Reinhold Kliegl, and Niels Landwehr

We study the problem of identifying individuals based on their characteristic gaze patterns during reading of arbitrary text. The motivation for this problem is an unobtrusive biometric setting in which a user is observed during access to a document, but no specific challenge protocol requiring the user's time and attention is carried out. Existing models of individual differences in gaze control during reading are either based on simple aggregate features of eye movements, or rely on parametric density models to describe, for instance, saccade amplitudes or word fixation durations. We develop flexible semiparametric models of eye movements during reading in which densities are inferred under a Gaussian process prior centered at a parametric distribution family that is expected to approximate the true distribution

well. An empirical study on reading data from 251 individuals shows significant improvements over the state of the art.

[Sentiment Analysis] [ID:L23]

Inducing Domain-Specific Sentiment Lexicons from Unlabeled Corpora
William L. Hamilton, Kevin Clark, Jure Leskovec, and Dan Jurafsky

A word's sentiment depends on the domain in which it is used. Computational social science research thus requires sentiment lexicons that are specific to the domains being studied. We combine domain-specific word embeddings with a label propagation framework to induce accurate domain-specific sentiment lexicons using small sets of seed words, achieving state-of-the-art performance competitive with approaches that rely on hand-curated resources. Using our framework we perform two large-scale empirical studies to quantify the extent to which sentiment varies across time and between communities. We induce and release historical sentiment lexicons for 150 years of English and community-specific sentiment lexicons for 250 online communities from the social media forum Reddit. The historical lexicons show that more than 5% of sentiment-bearing (non-neutral) English words completely switched polarity during the last 150 years, and the community-specific lexicons highlight how sentiment varies drastically between different communities.

[Sentiment Analysis] [ID:L24]

Attention-based LSTM for Aspect-level Sentiment Classification
Yequan Wang, Minlie Huang, and Li Zhao

Aspect-level sentiment classification is a fine-grained task in sentiment analysis. Since it provides more complete and in-depth results, aspect-level sentiment analysis has received much attention these years. In this paper, we reveal that the sentiment polarity of a sentence is not only determined by the content but is also highly related to the concerned aspect. For instance, "The appetizers are ok, but the service is slow.", for aspect taste, the polarity is positive while for service, the polarity is negative. Therefore, it is worthwhile to explore the connection between an aspect and the content of a sentence. To this end, we propose an Attention-based Long Short-Term Memory Network for aspect-level sentiment classification. The attention mechanism can concentrate on different parts of a sentence when different aspects are taken as input. We experiment on the SemEval 2014 dataset and results show that our model achieves state-of-the-art performance on aspect-level sentiment classification.

[Sentiment Analysis] [ID:L25]

Recursive Neural Conditional Random Fields for Aspect-based Sentiment Analysis
Wenya Wang, Sinno Jialin Pan, Daniel Dahlmeier, and Xiaokui Xiao

In aspect-based sentiment analysis, extracting aspect terms along with the opinions being expressed from user-generated content is one of the most important subtasks. Previous studies have shown that exploiting connections between aspect and opinion terms is promising for this task. In this paper, we propose a novel joint model that integrates recursive neural networks and conditional random fields into a unified framework for explicit aspect and opinion terms co-extraction. The proposed model learns high-level discriminative features and double propagate information between aspect and opinion terms, simultaneously. Moreover, it is flexible to incorporate hand-crafted features into the proposed model to further boost its information extraction performance. Experimental results on the dataset from SemEval Challenge 2014 show the superiority of our proposed model over several baseline methods as well as the winning systems of the challenge.

[Sentiment Analysis] [ID:L26]

Extracting Aspect Specific Opinion Expressions
Abhishek Laddha and Arjun Mukherjee

Opinionated expression extraction is a central problem in fine-grained sentiment analysis. Most existing works focus on either generic subjective expression or aspect expression extraction. However, in opinion mining, it is often desirable to mine the aspect specific opinion expressions (or aspectsentiment phrases) containing both the aspect and the opinion. This paper proposes a hybrid generative-discriminative framework for extracting such expressions. The hybrid model consists of (i) an unsupervised generative component for modeling the semantic coherence of terms (words/phrases) based on their collocations across different documents, and (ii) a supervised discriminative sequence modeling component for opinion phrase extraction. Experimental results using Amazon.com reviews demonstrate the effectiveness of the approach that significantly outperforms several state-of-the-art baselines.

[Sentiment Analysis] [ID:L27]

Emotion Distribution Learning from Texts
Deyu ZHOU, Xuan Zhang, Yin Zhou, Quan Zhao, and Xin Geng

The advent of social media and its prosperity enable users to share their opinions and views. Understanding users' emotional states might provide the potential to create new business opportunities. Automatically identifying users' emotional states from their texts and classifying emotions into finite categories such as joy, anger, disgust, etc., can be considered as a text classification problem. However, it introduces a challenging learning scenario where multiple emotions with different intensities are often found in a single sentence. Moreover, some emotions co-occur more often while other emotions rarely co-exist. In this paper, we propose a novel approach based on emotion distribution learning in order to address the aforementioned issues. The key idea is to learn a mapping function from sentences to their emotion distributions describing multiple emotions and their respective intensities. Moreover, the relations of emotions are captured based on the Plutchik's wheel of emotions and are subsequently incorporated into the learning algorithm in order to improve the accuracy of emotion detection. Experimental results show that the proposed approach can effectively deal with the emotion distribution detection problem and perform remarkably better than both the state-of-the-art emotion detection method and multi-label learning methods.

[Semantics] [ID:L28]

Building an Evaluation Scale using Item Response Theory
John Lalor and Hao Wu

Evaluation of NLP methods requires testing against a previously vetted gold-standard test set and reporting standard metrics (accuracy/precision/recall/F1). The current assumption is that all items in a given test set are equal with regards to difficulty and discriminating power. We propose Item Response Theory (IRT) from psychometrics as an alternative means for gold-standard test-set generation and NLP system evaluation. IRT is able to describe characteristics of individual items - their difficulty and discriminating power - and can account for these characteristics in its estimation of human intelligence or ability for an NLP task. In this paper, we demonstrate IRT by generating a gold-standard test set for Recognizing Textual Entailment. By collecting a large number of human responses and fitting our IRT model, we show that our IRT model compares NLP systems with the performance in a human population and is able to provide more insight into system performance than standard evaluation metrics. We show that a high accuracy score does not always imply a high IRT score, which depends on the item characteristics and the response pattern.

[Semantics] [ID:L29]

WordRank: Learning Word Embeddings via Robust Ranking
Shihao Ji, Hyokun Yun, Pinar Yanardag, Shin Matsushima, and S. V. N. Vishwanathan

Embedding words in a vector space has gained a lot of attention in recent years. While state-of-the-art methods provide efficient computation of word similarities via a low-dimensional matrix embedding, their motivation is often left unclear. In this paper, we argue that word embedding can be naturally viewed as a ranking problem due to the ranking nature of the evaluation metrics. Then, based on this insight, we propose a novel framework WordRank that efficiently estimates word representations via robust ranking, in which the attention mechanism and robustness to noise are readily achieved via the DCG-like ranking losses. The performance of WordRank is measured in word similarity and word analogy benchmarks, and the results are compared to the state-of-the-art word embedding techniques. Our algorithm is very competitive to the state-of-the-arts on large corpora, while outperforms them by a significant margin when the training set is limited (i.e., sparse and noisy). With 17 million tokens, WordRank performs almost as well as existing methods using 7.2 billion tokens on a popular word similarity benchmark. Our multi-node distributed implementation of WordRank is publicly available for general usage.

[Semantics] [ID:L30]

Exploring Semantic Representation in Brain Activity Using Word Embeddings
Yu-Ping Ruan, Zhen-Hua Ling, and Yu Hu

In this paper, we utilize distributed word representations (i.e., word embeddings) to analyse the representation of semantics in brain activity. The brain activity data were recorded using functional magnetic resonance imaging (fMRI) when subjects were viewing words. First, we analysed the functional selectivity of different cortex areas by calculating the correlations between neural responses and several types of word representations, including skip-gram word embeddings, visual semantic vectors, and primary visual features. The results demonstrated consistency with existing neuroscientific knowledge. Second, we utilized behavioural data as the semantic ground truth to measure their relevance with brain activity. A method to estimate word embeddings under the constraints of brain activity similarities is further proposed based on the semantic word embedding (SWE) model. The experimental results show that the brain activity data are significantly correlated with the behavioural data of human judgements on semantic similarity. The correlations between the estimated word embeddings and the semantic ground truth can be effectively improved after integrating the brain activity data for learning, which implies that semantic patterns in neural representations may exist that have not been fully captured by state-of-the-art word embeddings derived from text corpora.

[Semantics] [ID:L31]

AMR Parsing with an Incremental Joint Model

Junsheng Zhou, Feiyu Xu, Hans Uszkoreit, Weiguang QU, Ran Li, and Yanhui Gu

To alleviate the error propagation in the traditional pipelined models for Abstract Meaning Representation (AMR) parsing, we formulate AMR parsing as a joint task that performs the two subtasks: concept identification and relation identification simultaneously. To this end, we first develop a novel component-wise beam search algorithm for relation identification in an incremental fashion, and then in-corporate the decoder into a unified frame-work based on multiple-beam search, which allows for the bi-directional information flow between the two subtasks in a single incremental model. Experiments on the public datasets demonstrate that our joint model significantly outperforms the previous pipelined counterparts, and also achieves better or comparable performance than other approaches to AMR parsing, without utilizing external semantic resources.

[Social Media & Computational Social Science] [ID:L32]

Identifying Dogmatism in Social Media: Signals and Models

Ethan Fast and Eric Horvitz

We explore linguistic and behavioral features of dogmatism in social media and construct statistical models that can identify dogmatic comments. Our model is based on a corpus of Reddit posts, collected across a diverse set of conversational topics and annotated via paid crowdsourcing. We operationalize key aspects of dogmatism described by existing psychology theories (such as over-confidence), finding they have predictive power. We also find evidence for new signals of dogmatism, such as the tendency of dogmatic posts to refrain from signaling cognitive processes. When we use our predictive model to analyze millions of other Reddit posts, we find evidence that suggests dogmatism is a deeper personality trait, present for dogmatic users across many different domains, and that users who engage on dogmatic comments tend to show increases in dogmatic posts themselves.

[Social Media & Computational Social Science] [ID:L33]

Enhanced Personalized Search using Social Data

Dong Zhou, Séamus Lawless, Xuan Wu, Wenyu Zhao, and Jianxun Liu

Search personalization that considers the social dimension of the web has attracted a significant volume of research in recent years. A user profile is usually needed to represent a user's interests in order to tailor future searches. Previous research has typically constructed a profile solely from a user's usage in-formation. When the user has only limited activities in the system, the effect of the user profile on search is also constrained. This research addresses the setting where a user has only a limited amount of usage information. We build enhanced user profiles from a set of annotations and resources that users have marked, together with an external knowledge base constructed according to usage histories. We present two probabilistic latent topic models to simultaneously incorporate social annotations, documents and the external knowledge base. Our web search strategy is achieved using personalized social query expansion. We introduce a topical query expansion model to enhance the search by utilizing individual user profiles. The proposed approaches have been intensively evaluated on a large pub-lic social annotation dataset. Results show that our models significantly outperformed existing personalized query expansion methods which use user profiles solely built from past usage in-formation in personalized search.

[Syntax & Morphology] [ID:L34]

Effective Greedy Inference for Graph-based Non-Projective Dependency Parsing

Ilan Tchernowitz, Liron Yedidsion, and Roi Reichart

Exact inference in high-order graph-based non-porjective dependency parsing is intractable. Hence, sophisticated approximation techniques based on algorithms such as belief propagation and dual decomposition have been employed. In contrast, we propose a simple greedy search approximation for this problem which is very intuitive and easy to implement. We implement the algorithm within the second-order TurboParser and experiment with the datasets of the CoNLL 2006 and 2007 shared task on multilingual dependency parsing. Our algorithm improves the run time of the parser by a factor of 1.43 while losing 1% in UAS on average across languages. Moreover, an ensemble method exploiting the joint power of the parsers, achieves an average UAS 0.27% higher than the TurboParser.

[Syntax & Morphology] [ID:L35]

Generating Abbreviations for Chinese Named Entities Using Recurrent Neural Network with Dynamic Dictionary

Qi Zhang, Jin Qian, Ya Guo, Yaqian Zhou, and Xuanjing Huang

Chinese named entities occur frequently in formal and informal environments. Various approaches have been formalized the problem as a sequence labelling task and utilize a character-based methodology, in which character is treated as the basic classification unit. One of the main drawbacks of these methods is that some of the generated abbreviations may not follow the conventional wisdom of Chinese. To

address this problem, we propose a novel neural network architecture to perform task. It combines recurrent neural network (RNN) with an architecture determining whether a given sequence of characters can be a word or not. For demonstrating the effectiveness of the proposed method, we evaluate it on Chinese named entity generation and opinion target extraction tasks. Experimental results show that the proposed method can achieve better performance than state-of-the-art methods.

[Syntax & Morphology] [ID:L36]

Neural Network for Heterogeneous Annotations
Hongshen Chen, Yue Zhang, and Qun Liu

Multiple treebanks annotated under heterogeneous standards give rise to the research question of best utilizing multiple resources for improving statistical models. Prior research has focused on discrete models, leveraging stacking and multi-view learning to address the problem. In this paper, we empirically investigate heterogeneous annotations using neural network models, building a neural network counterpart to discrete stacking and multiview learning, respectively, finding that neural models have their unique advantages thanks to the freedom from manual feature engineering. Neural model achieves not only better accuracy improvements, but also an order of magnitude faster speed compared to its discrete baseline, adding little time cost compared to a neural model trained on a single treebank.

[Syntax & Morphology] [ID:L37]

LAMB: A Good Shepherd of Morphologically Rich Languages
Sebastian Ebert, Thomas Müller, and Hinrich Schütze

This paper introduces STEM and LAMB, embeddings trained for stems and lemmata instead of for surface forms. For morphologically rich languages, they perform significantly better than standard embeddings on word similarity and polarity evaluations. On a new WordNet-based evaluation, STEM and LAMB are up to 50% better than standard embeddings. We show that both embeddings have high quality even for small dimensionality and training corpora.

[Syntax & Morphology] [ID:L38]

Fast Coupled Sequence Labeling on Heterogeneous Annotations via Context-aware Pruning
Zhenghua Li, Jiayuan Chao, Min Zhang, and Jiwen Yang

The recently proposed coupled sequence labeling is shown to be able to effectively exploit multiple labeled data with heterogeneous annotations but suffer from severe inefficiency problem due to the large bundled tag space (Li et al., 2015). In their case study of part-of-speech (POS) tagging, Li et al. (2015) manually design context-free tag-to-tag mapping rules with a lot of effort to reduce the tag space. This paper proposes a context-aware pruning approach that performs token-wise constraints on the tag space based on contextual evidences, making the coupled approach efficient enough to be applied to the more complex task of joint word segmentation (WS) and POS tagging for the first time. Experiments show that using the large-scale People Daily as auxiliary heterogeneous data, the coupled approach can improve F-score by 95.55-94.88=0.67% on WS, and by 90.58-89.49=1.09% on joint WS and POS on Penn Chinese Treebank.

[Syntax & Morphology] [ID:L39]

Unsupervised Neural Dependency Parsing
Yong Jiang, Wenjuan Han, and Kewei Tu

Unsupervised dependency parsing aims to learn a dependency grammar from text annotated with only POS tags. Various features and inductive biases are often used to incorporate prior knowledge into learning. One useful type of prior information is that there exist correlations between the parameters of grammar rules involving different POS tags. Previous work employed manually designed features or special prior distributions to encode such information. In this paper, we propose a novel approach to unsupervised dependency parsing that uses a neural model to predict grammar rule probabilities based on distributed representation of POS tags. The distributed representation is automatically learned from data and captures the correlations between POS tags. Our experiments show that our approach outperforms previous approaches utilizing POS correlations and is competitive with recent state-of-the-art approaches on nine different languages.

[Summarization] [ID:L40]

Generating Coherent Summaries of Scientific Articles Using Coherence Patterns
Daraksha Parveen, Mohsen Mesgar, and Michael Strube

Previous work on automatic summarization does not thoroughly consider coherence while generating the summary. We introduce a graph-based approach to summarize scientific articles. We employ coherence

patterns to ensure that the generated summaries are coherent. The novelty of our model is twofold: we mine coherence patterns in a corpus of abstracts, and we propose a method to combine coherence, importance and non-redundancy to generate the summary. We optimize these factors simultaneously using Mixed Integer Pro- gramming. Our approach significantly outperforms baseline and state-of-the-art systems in terms of coherence (summary coherence assessment) and relevance (ROUGE scores).

[Summarization] [ID:L41]

News Stream Summarization using Burst Information Networks

Tao Ge, Lei Cui, Baobao Chang, Sujian Li, Ming Zhou, and Zhifang Sui

This paper studies summarizing key information from news streams. We propose simple yet effective models to solve the problem based on a novel and promising representation of text streams – Burst Information Networks (BINets). A BINet can be aware of redundant information, allows global analysis of a text stream, and can be efficiently built and dynamically updated, which perfectly fits the demands of text stream summarization. Extensive experiments show that the BINet-based approaches are not only efficient and can be used in a real-time online summarization setting, but also can generate high-quality summaries, outperforming the state-of-the-art approach.

[Text Mining & Applications] [ID:L42]

Rationale-Augmented Convolutional Neural Networks for Text Classification

Ye Zhang, Iain Marshall, and Byron C. Wallace

We present a new Convolutional Neural Network (CNN) model for text classification that jointly exploits labels on documents and their constituent sentences. Specifically, we consider scenarios in which annotators explicitly mark sentences (or snippets) that support their overall document categorization, i.e., they provide rationales. Our model exploits such supervision via a hierarchical approach in which each document is represented by a linear combination of the vector representations of its component sentences. We propose a sentence-level convolutional model that estimates the probability that a given sentence is a rationale, and we then scale the contribution of each sentence to the aggregate document representation in proportion to these estimates. Experiments on five classification datasets that have document labels and associated rationales demonstrate that our approach consistently outperforms strong baselines. Moreover, our model naturally provides explanations for its predictions.

[Text Mining & Applications] [ID:L43]

Transferring User Interests Across Websites with Unstructured Text for Cold-Start Recommendation

Yu-Yang Huang and Shou-De Lin

In this work, we investigate the possibility of cross-website transfer learning for tackling the cold-start problem. To address the cold-start issues commonly present in a collaborative filtering (CF) system, most existing cross-domain CF models require auxiliary rating data from another domain; nevertheless, under the cross-website scenario, such data is often unobtainable. Therefore, we propose the nearest-neighbor transfer matrix factorization (NT-MF) model, where a topic model is applied to the unstructured user-generated content in the source domain, and the similarity between users in the latent topic space is utilized to guide the target-domain CF model. Specifically, the latent factors of the nearest-neighbors are regarded as a set of pseudo observations, which can be used to estimate the unknown parameters in the model. Improvement over previous methods, especially for the cold-start users, is demonstrated with experiments on a real-world cross-website dataset.

[Text Mining & Applications] [ID:L44]

Speculation and Negation Scope Detection via Convolutional Neural Networks

Zhong Qian, Peifeng Li, Qiaoming Zhu, Guodong Zhou, Zhunchen Luo, and Wei Luo

Speculation and negation are important in-formation to identify text factuality. In this paper, we propose a Convolutional Neural Network (CNN)-based model with probabil-istic weighted average pooling to address speculation and negation scope detection. In particular, our CNN-based model extracts those meaningful features from various syn-tactic paths between the cues and the candi-date tokens in both constituency and depen-dency parse trees. Evaluation on BioScope shows that our CNN-based model significant-ly outperforms the state-of-the-art systems on Abstracts, a sub-corpus in BioScope, and achieves comparable performances on Clini-cal Records, another sub-corpus in BioScope.

[Text Mining & Applications] [ID:L45]

Analyzing Linguistic Knowledge in Sequential Model of Sentence

Peng Qian, Xipeng Qiu, and Xuanjing Huang

Sentence modelling is a fundamental topic in computational linguistics. Recently, deep learning-based sequential models of sentence, such as recurrent neural network, have proved to be effective in dealing

with the non-sequential properties of human language. However, little is known about how a recurrent neural network captures linguistic knowledge. Here we propose to correlate the neuron activation pattern of a LSTM language model with rich language features at sequential, lexical and compositional level. Qualitative visualization as well as quantitative analysis under multilingual perspective reveals the effectiveness of gate neurons and indicates that LSTM learns to allow different neurons selectively respond to linguistic knowledge at different levels. Cross-language evidence shows that the model captures different aspects of linguistic properties for different languages due to the variance of syntactic complexity. Additionally, we analyze the influence of modelling strategy on linguistic knowledge encoded implicitly in different sequential models.

[Text Mining & Applications] [ID:L46]

Keyphrase Extraction Using Deep Recurrent Neural Networks on Twitter

Qi Zhang, Yang Wang, Yeyun Gong, and Xuanjing Huang

Keyphrases can provide highly condensed and valuable information that allows users to quickly acquire the main ideas. The task of automatically extracting them have received considerable attention in recent decades. Different from previous studies, which are usually focused on automatically extracting keyphrases from documents or articles, in this study, we considered the problem of automatically extracting keyphrases from tweets. Because of the length limitations of Twitter-like sites, the performances of existing methods usually drop sharply. We proposed a novel deep recurrent neural network (RNN) model to combine keywords and context information to perform this problem. To evaluate the proposed method, we also constructed a large-scale dataset collected from Twitter. The experimental results showed that the proposed method performs significantly better than previous methods.

[Text Mining & Applications] [ID:L47]

Solving and Generating Chinese Character Riddles

Chuanqi Tan, Furu Wei, Li Dong, Weifeng Lv, and Ming Zhou

Chinese character riddle is a riddle game in which the riddle solution is a single Chinese character. It is closely connected with the shape, pronunciation or meaning of Chinese characters. The riddle description (sentence) is usually composed of phrases with rich linguistic phenomena (such as pun, simile, and metaphor), which are associated to different parts (namely radicals) of the solution character. In this paper, we propose a statistical framework to solve and generate Chinese character riddles. Specifically, we learn the alignments and rules to identify the metaphors between phrases in riddles and radicals in characters. Then, in the solving phase, we utilize a dynamic programming method to combine the identified metaphors to obtain candidate solutions. In the riddle generation phase, we use a template-based method and a replacement-based method to obtain candidate riddle descriptions. We then use Ranking SVM to rerank the candidates both in the solving and generation process. Experimental results in the solving task show that the proposed method outperform baseline methods. We also get very promising results in the generation task according to human judges.

[Text Mining & Applications] [ID:L48]

Structured prediction models for RNN based sequence labeling in clinical text

Abhyuday Jagannatha

Sequence labeling is a widely used method for named entity recognition and information extraction from unstructured natural language data. In the clinical domain one major application of sequence labeling involves extraction of relevant entities such as medication, indication, and side-effects from Electronic Health Record Narratives. Sequence labeling in this domain presents its own set of challenges and objectives. In this work we experiment with Conditional Random Field based structured learning models with Recurrent Neural Networks. We extend the previously studied CRF-LSTM model with explicit modeling of pairwise potentials. We also propose an approximate version of skip-chain CRF inference with RNN potentials. We use these methods for structured prediction in order to improve the exact phrase detection of clinical entities.

[Text Mining & Applications] [ID:L49]

Learning to Represent Review with Tensor Decomposition for Spam Detection

Xuepeng Wang, Kang Liu, Shizhu He, and Jun Zhao

Review spam detection is a key task in opinion mining. To accomplish this type of detection, previous work has focused mainly on effectively representing fake and non-fake reviews with discriminative features, which are discovered or elaborately designed by experts or developers. This paper proposes a novel review spam detection method that learns the representation of reviews automatically instead of heavily relying on experts' knowledge in a data-driven manner. More specifically, according to 11 relations (generated automatically from two basic patterns) between reviewers and products, we employ tensor decomposition to learn the embeddings of the reviewers and products in a vector space. We collect relations between any two entities (reviewers and products), which results in much useful and

global information. We concatenate the review text, the embeddings of the reviewer and the reviewed product as the representation of a review. Based on such representations, the classifier could identify the opinion spam more precisely. Experimental results on an open Yelp dataset show that our method could effectively enhance the spam detection accuracy compared with the state-of-the-art methods.

[Text Mining & Applications] [ID:L50]

Stance Detection with Bidirectional Conditional Encoding

Isabelle Augenstein, Tim Rocktäschel, Andreas Vlachos, and Kalina Bontcheva

Stance detection is the task of classifying the attitude expressed in a text towards a target such as Hillary Clinton to be "positive", "negative" or "neutral". Previous work has assumed that either the target is mentioned in the text or that training data for every target is given. This paper considers the more challenging version of this task, where targets are not always mentioned and no training data is available for the test targets. We experiment with conditional LSTM encoding, which builds a representation of the tweet that is dependent on the target, and demonstrate that it outperforms encoding the tweet and the target independently. Performance is improved further when the conditional model is augmented with bidirectional encoding. We evaluate our approach on the SemEval 2016 Task 6 Twitter Stance Detection corpus achieving performance second best only to a system trained on semi-automatically labelled tweets for the test target. When such weak supervision is added, our approach achieves state-of-the-art results.

[Information Extraction] [ID:S01]

Modeling Skip-Grams for Event Detection with Convolutional Neural Networks

Thien Huu Nguyen and Ralph Grishman

Convolutional neural networks (CNN) have achieved the top performance for event detection due to their capacity to induce the underlying structures of the k-grams in the sentences. However, the current CNN-based event detectors only model the consecutive k-grams and ignore the non-consecutive k-grams that might involve important structures for event detection. In this work, we propose to improve the current CNN models for ED by introducing the non-consecutive convolution. Our systematic evaluation on both the general setting and the domain adaptation setting demonstrates the effectiveness of the non-consecutive CNN model, leading to the significant performance improvement over the current state-of-the-art systems.

[Information Extraction] [ID:S02]

Porting an Open Information Extraction System from English to German

Tobias Falke, Gabriel Stanovsky, Iryna Gurevych, and Ido Dagan

Many downstream NLP tasks can benefit from Open Information Extraction (Open IE) as a semantic representation. While Open IE systems are available for English, many other languages lack such tools. In this paper, we present a straightforward approach for adapting PropS, a rule-based predicate-argument analysis for English, to a new language, German. With this approach, we quickly obtain an Open IE system for German covering 89% of the English rule set. It yields 1.6 n-ary extractions per sentence at 60% precision, making it comparable to systems for English and readily usable in downstream applications.

[Information Extraction] [ID:S03]

Named Entity Recognition for Novel Types by Transfer Learning

Lizhen Qu, Gabriela Ferraro, Liyuan Zhou, Weiwei Hou, and Timothy Baldwin

In named entity recognition, we often don't have a large in-domain training corpus or a knowledge base with adequate coverage to train a model directly. In this paper, we propose a method where, given training data in a related domain with similar (but not identical) named entity (NE) types and a small amount of in-domain training data, we use transfer learning to learn a domain-specific NE model. That is, the novelty in the task setup is that we assume not just domain mismatch, but also label mismatch.

[Information Extraction] [ID:S04]

Extracting Subevents via an Effective Two-phase Approach

Allison Badgett and Ruihong Huang

We present our pilot research on automatically extracting subevents from a domain-specific corpus, focusing on the type of subevents that describe physical actions composing an event. We decompose the challenging problem and propose a two-phase approach that effectively captures sentential and local cues that describe subevents. We extracted a rich set of over 600 novel subevent phrases. Evaluation shows the automatically learned subevents help to discover 10% additional main events (of which the learned subevents are a part) and improve event detection performance.

[Language and Vision] [ID:S05]

Gaussian Visual-Linguistic Embedding for Zero-Shot Recognition
Tanmoy Mukherjee and Timothy Hospedales

An exciting outcome of research at the intersection of language and vision is that of zero-shot learning (ZSL). ZSL promises to scale visual recognition by borrowing distributed semantic models learned from linguistic corpora and turning them into visual recognition models. However the popular word-vector DSM embeddings are relatively impoverished in their expressivity as they model each word as a single vector point. In this paper we explore word distribution embeddings for ZSL. We present a visual-linguistic mapping for ZSL in the case where words and visual categories are both represented by distributions. Experiments show improved results on ZSL benchmarks due to this better exploiting of intra-concept variability in each modality

[Language and Vision] [ID:S06]
Question Relevance in VQA: Identifying Non-Visual And False-Premise Questions
Arijit Ray, Gordon Christie, Mohit Bansal, Dhruv Batra, and Devi Parikh

Visual Question Answering (VQA) is the task of answering natural-language questions about images. We introduce the novel problem of determining the relevance of questions to images in VQA. Current VQA models do not reason about whether a question is even related to the given image (e.g., What is the capital of Argentina?) or if it requires information from external resources to answer correctly. This can break the continuity of a dialogue in human-machine interaction. Our approaches for determining relevance are composed of two stages. Given an image and a question, (1) we first determine whether the question is visual or not, (2) if visual, we determine whether the question is relevant to the given image or not. Our approaches, based on LSTM-RNNs, VQA model uncertainty, and caption-question similarity, are able to outperform strong baselines on both relevance tasks. We also present human studies showing that VQA models augmented with such question relevance reasoning are perceived as more intelligent, reasonable, and human-like.

[Language and Vision] [ID:S07]
Sort Story: Sorting Jumbled Images and Captions into Stories
Harsh Agrawal, Arjun Chandrasekaran, Dhruv Batra, Devi Parikh, and Mohit Bansal

Temporal common sense has applications in AI tasks such as QA, multi-document summarization, and human-AI communication. We propose the task of sequencing – given a jumbled set of aligned image-caption pairs that belong to a story, the task is to sort them such that the output sequence forms a coherent story. We present multiple approaches, via unary (position) and pairwise (order) predictions, and their ensemble-based combinations, achieving strong results on this task. We use both text-based and image-based features, which depict complementary improvements. Using qualitative examples, we demonstrate that our models have learnt interesting aspects of temporal common sense.

[Language and Vision] [ID:S08]
Human Attention in Visual Question Answering: Do Humans and Deep Networks look at the same regions?
Abhishek Das, Harsh Agrawal, Larry Zitnick, Devi Parikh, and Dhruv Batra

We conduct large-scale studies on 'human attention' in Visual Question Answering (VQA) to understand where humans choose to look to answer questions about images. We design and test multiple game-inspired novel attention-annotation interfaces that require the subject to sharpen regions of a blurred image to answer a question. Thus, we introduce the VQA-HAT (Human ATtention) dataset. We evaluate attention maps generated by state-of-the-art VQA models against human attention both qualitatively (via visualizations) and quantitatively (via rank-order correlation). Overall, our experiments show that current VQA attention models do not seem to be looking at the same regions as humans.

[Machine Learning] [ID:S09]
Recurrent Residual Learning for Sequence Classification
Yiren Wang and Fei Tian

In this paper, we explore the possibility of leveraging Residual Networks (ResNet), a powerful structure in constructing extremely deep neural network for image understanding, to improve recurrent neural networks (RNN) for modeling sequential data. We show that for sequence classification tasks, incorporating residual connections into recurrent structures yields similar accuracy to Long Short Term Memory (LSTM) RNN with much fewer model parameters. In addition, we propose two novel models which combine the best of both residual learning and LSTM. Experiments show that the new models significantly outperform LSTM.

[Machine Learning] [ID:S10]
Richer Interpolative Smoothing Based on Modified Kneser-Ney Language Modeling
Ehsan Shareghi, Trevor Cohn, and Gholamreza Haffari

In this work we present a generalisation of the Modified Kneser-Ney interpolative smoothing for richer smoothing via additional discount parameters. We provide mathematical underpinning for the estimator of the new discount parameters, and showcase the utility of our rich MKN language models on several European languages. We further explore the interdependency among the training data size, language model order, and number of discount parameters. Our empirical results illustrate that larger number of discount parameters, i) allows for better allocation of mass in the smoothing process, particularly on small data regime where statistical sparsity is sever, and ii) leads to significant reduction in perplexity, particularly for out-of-domain test sets which introduce higher ratio of out-of-vocabulary words.

[Machine Learning] [ID:S11]

A General Regularization Framework for Domain Adaptation

Wei Lu, Hai Leong Chieu, and Jonathan Löfgren

We propose a domain adaptation framework, and formally prove that it generalizes the feature augmentation technique in (Daume III, 2007) and the multi-task regularization framework in (Evgeniou and Pontil, 2004). We show that our framework is strictly more general than these approaches and allows practitioners to tune hyper-parameters to encourage transfer between close domains and avoid negative transfer between distant ones.

[Machine Translation] [ID:S12]

Coverage Embedding Models for Neural Machine Translation

Haitao Mi, Baskaran Sankaran, Zhiguo Wang, and Abe Ittycheriah

In this paper, we enhance the attention-based neural machine translation (NMT) by adding explicit coverage embedding models to alleviate issues of repeating and dropping translations in NMT. For each source word, our model starts with a full coverage embedding vector to track the coverage status, and then keeps updating it with neural networks as the translation goes. Experiments on the large-scale Chinese-to-English task show that our enhanced model improves the translation quality significantly on various test sets over the strong large vocabulary NMT system.

[Syntax & Morphology] [ID:S13]

Neural Morphological Analysis: Encoding-Decoding Canonical Segments

Katharina Kann, Ryan Cotterell, and Hinrich Schütze

Canonical morphological segmentation aims to divide words into a sequence of standardized segments. In this work, we propose a character-based neural encoder-decoder model for this task. Additionally, we extend our model to include morpheme-level and lexical information through a neural reranker. We set the new state of the art for the task improving previous results by up to 21% accuracy. Our experiments cover three languages: English, German and Indonesian.

[Syntax & Morphology] [ID:S14]

Exploiting Mutual Benefits between Syntax and Semantic Roles using Neural Network

Peng Shi, Zhiyang Teng, and Yue Zhang

We investigate mutual benefits between syntax and semantic roles using neural network models, by studying a parsing–>SRL pipeline, a SRL–>parsing pipeline, and a simple joint model by embedding sharing. The integration of syntactic and semantic features gives promising results in a Chinese Semantic Tree-bank, demonstrating large potentials of neural models for joint parsing and semantic role labeling.

[Semantics] [ID:S15]

The Effects of Data Size and Frequency Range on Distributional Semantic Models

Magnus Sahlgren and Alessandro Lenci

This paper investigates the effects of data size and frequency range on Distributional Semantic Models. We compare the performance of a number of representative models for several test settings over data of varying sizes, and over test items of various frequency. Our results show that neural network-based models underperform when the data is small, and that the most reliable model over data of varying sizes and frequency ranges is the inverted factorized model.

[Semantics] [ID:S16]

Multi-Granularity Chinese Word Embedding

Rongchao Yin, Quan Wang, Peng Li, Rui Li, and Bin Wang

This paper considers the problem of learning Chinese word embeddings. In contrast to English, a Chinese word is usually composed of characters, and most of the characters themselves can be further

divided into components such as radicals. While characters and radicals contain rich information and are capable of indicating semantic meanings of words, they have not been fully exploited by existing word embedding methods. In this work, we propose multi-granularity embedding (MGE) for Chinese words. The key idea is to make full use of such word-character-radical composition, and enrich word embeddings by further incorporating finer-grained semantics from characters and radicals. Quantitative evaluation demonstrates the superiority of MGE in word similarity computation and analogical reasoning. Qualitative analysis further shows its capability to identify finer-grained semantic meanings of words.

[Semantics] [ID:S17]

Numerically Grounded Language Models for Semantic Error Correction

Georgios Spithourakis, Isabelle Augenstein, and Sebastian Riedel

Semantic error detection and correction is an important task for applications such as fact checking, speech-to-text or grammatical error correction. Current approaches generally focus on relatively shallow semantics and do not account for numeric quantities. Our approach uses language models grounded on numbers within the text. Such groundings are easily achieved for recurrent neural language model architectures, which can be further conditioned on incomplete background knowledge bases. Our evaluation on clinical reports shows that numerical grounding improves perplexity by 33% and F1 for semantic error correction by 5 points when compared to ungrounded approaches. Conditioning on a knowledge base yields further improvements.

[Semantics] [ID:S18]

Towards Semi-Automatic Generation of Proposition Banks for Low-Resource Languages

Alan Akbik and Yunyao Li

Annotation projection based on parallel corpora has shown great promise in inexpensively creating Proposition Banks for languages for which high-quality parallel corpora and syntactic parsers are available. In this paper, we conduct an experimental study where we apply this approach to three languages that lack such resources: Tamil, Bengali and Malayalam. We find an average quality difference of 6 to 20 absolute F-measure points vis-a-vis high-resource languages, which indicates that annotation projection alone is insufficient in low-resource scenarios. Based on these results, we explore the possibility of using annotation projection as a starting point for inexpensive data curation involving both experts and non-experts. We give an outline of what such a process may look like and present an initial study to discuss its potential and challenges.

[Sentiment Analysis] [ID:S19]

A Hierarchical Model of Reviews for Aspect-based Sentiment Analysis

Sebastian Ruder, Parsa Ghaffari, and John G. Breslin

Opinion mining from customer reviews has become pervasive in recent years. Sentences in reviews, however, are usually classified independently, even though they form part of a review's argumentative structure. Intuitively, sentences in a review build and elaborate upon each other; knowledge of the review structure and sentential context should thus inform the classification of each sentence. We demonstrate this hypothesis for the task of aspect-based sentiment analysis by modeling the interdependencies of sentences in a review with a hierarchical bidirectional LSTM. We show that the hierarchical model outperforms two non-hierarchical baselines, obtains results competitive with the state-of-the-art, and outperforms the state-of-the-art on five multilingual, multi-domain datasets without any hand-engineered features or external resources.

[Sentiment Analysis] [ID:S20]

Are Word Embedding-based Features Useful for Sarcasm Detection?

Aditya Joshi, Vaibhav Tripathi, Kevin Patel, Pushpak Bhattacharyya, and Mark Carman

This paper makes a simple increment to the state-of- the-art in sarcasm detection research. Existing approaches are unable to capture subtle forms of con- text incongruity which lies at the heart of sarcasm. We explore if prior work can be enhanced using se- mantic similarity/discordance between word embeddings. We augment word embedding-based features to four feature sets reported in the past. We experiment with four types of word embeddings, and observe an improvement in sarcasm detection, irrespective of the word embedding used or the original feature set to which our features are augmented. For example, this augmentation results in an improve- ment in F-score of around 4% for three out of these four feature sets, and a minor degradation in case of the fourth, when word2vec embeddings are used. Fi- nally, a comparison of the four embeddings shows that word2vec and dependency weight-based fea- tures outperform LSA and GloVe, in terms of their benefit to sarcasm detection.

[Sentiment Analysis] [ID:S21]

Weakly Supervised Tweet Stance Classification by Relational Bootstrapping
Javid Ebrahimi, Dejing Dou, and Daniel Lowd

Supervised stance classification, in such domains as Congressional debates and online forums, has been a topic of interest in the past decade. Approaches have evolved from text classification to structured output prediction, including collective classification and sequence labeling. In this work, we investigate collective classification of stances on Twitter, using hinge-loss Markov random fields (HL-MRFs). Given the graph of all posts, users, and their relationships, we constrain the predicted post labels and latent user labels to correspond with the network structure. We focus on a weakly supervised setting, in which only a small set of hashtags or phrases is labeled. Using our relational approach, we are able to go beyond the stance-indicative patterns and harvest more stance-indicative tweets, which can also be used to train any linear text classifier when the network structure is not available or is costly.

[Social Media & Computational Social Science] [ID:S22]
The Gun Violence Database: A new task and data set for NLP
Ellie Pavlick, Heng Ji, Xiaoman Pan, and Chris Callison-Burch

We argue that NLP researchers are especially well-positioned to contribute to the national discussion about gun violence. Reasoning about the causes and outcomes of gun violence is typically dominated by politics and emotion, and data-driven research on the topic is stymied by a shortage of data and a lack of federal funding. However, data abounds in the form of unstructured text from news articles across the country. This is an ideal application of NLP technologies, such as relation extraction, coreference resolution, and event detection. We introduce a new and growing dataset, the Gun Violence Database, in order to facilitate the adaptation of current NLP technologies to the domain of gun violence, thus enabling better social science research on this important and under-resourced problem.

[Social Media & Computational Social Science] [ID:S23]
Fluency detection on communication networks
Tom Lippincott and Benjamin Van Durme

When considering a social media corpus, we often have access to structural information about how messages are flowing between people or organizations. This information is particularly useful when the linguistic evidence is sparse, incomplete, or of dubious quality. In this paper we construct a simple model to leverage the structure of Twitter data to help determine the set of languages each user is fluent in. Our results demonstrate that imposing several intuitive constraints leads to improvements in performance and stability. We release the first annotated data set for exploring this task, and discuss how our approach may be extended to other applications.

[Social Media & Computational Social Science] [ID:S24]
Cultural Shift or Linguistic Drift? Comparing Two Computational Measures of Semantic Change
William L. Hamilton, Jure Leskovec, and Dan Jurafsky

Words shift in meaning for many reasons, including cultural factors like new technologies and regular linguistic processes like subjectification. Understanding the evolution of language and culture requires disentangling these underlying causes. Here we show how two different distributional measures can be used to detect two different types of semantic change. The first measure, which has been used in many previous works, analyzes global shifts in a word's distributional semantics; it is sensitive to changes due to regular processes of linguistic drift, such as the semantic generalization of promise ("I promise." -> "It promised to be exciting."). The second measure, which we develop here, focuses on local changes to a word's nearest semantic neighbors; it is more sensitive to cultural shifts, such as the change in the meaning of cell ("prison cell" ->"cell phone"). Comparing measurements made by these two methods allows researchers to determine whether changes are more cultural or linguistic in nature, a distinction that is essential to work in the digital humanities and historical linguistics.

[Social Media & Computational Social Science] [ID:S25]
Characterizing the Language of Online Communities and its Relation to Community Reception
Trang Tran and Mari Ostendorf

This work investigates style and topic aspects of language in online communities: looking at both utility as an identifier of the community and correlation with community reception of content. Style is characterized using hybrid word and part-of-speech tag n-gram language models, while topic is represented using Latent Dirichlet Allocation. Experiments with several Reddit forums show that style is a better indicator of community identity than topic, even for communities organized around specific topics. Further, there is a positive correlation between the community reception to a contribution and the style similarity to that community, but not so for topic similarity.

[Spoken Language Processing] [ID:S26]

Joint Transition-based Dependency Parsing and Disfluency Detection for Automatic Speech Recognition Texts

Masashi Yoshikawa, Hiroyuki Shindo, and Yuji Matsumoto

Joint dependency parsing with disfluency detection is an important task in speech language processing. Recent methods show high performance for this task, although most authors make the unrealistic assumption that input texts are transcribed by human annotators. In real-world applications, the input text is typically the output of an automatic speech recognition (ASR) system, which implies that the text contains not only disfluency noises but also recognition errors from the ASR system. In this work, we propose a parsing method that handles both disfluency and ASR errors using an incremental shift-reduce algorithm with several novel features suited to ASR output texts. Because the gold dependency information is usually annotated only on transcribed texts, we also introduce an alignment-based method for transferring the gold dependency annotation to the ASR output texts to construct training data for our parser. We conducted an experiment on the Switchboard corpus and show that our method outperforms conventional methods in terms of dependency parsing and disfluency detection.

[Spoken Language Processing] [ID:S27]

Real-Time Speech Emotion and Sentiment Recognition for Interactive Dialogue Systems

Dario Bertero, Farhad Bin Siddique, Chien-Sheng Wu, Yan Wan, Ricky Ho Yin Chan, and Pascale Fung

In this paper, we describe our approach of enabling an interactive dialogue system to recognize user emotion and sentiment in realtime. These modules allow otherwise conventional dialogue systems to have "empathy" and answer to the user while being aware of their emotion and intent. Emotion recognition from speech previously consists of feature engineering and machine learning where the first stage causes delay in decoding time. We describe a CNN model to extract emotion from raw speech input without feature engineering. This approach even achieves an impressive average of 65.7% accuracy on six emotion categories, a 4.5% improvement when compared to the conventional feature based SVM classification. A separate, CNN-based sentiment analysis module recognizes sentiments from speech recognition results, with 74.8 F-measure on human-machine dialogues when trained with out-of-domain data.

[Summarization] [ID:S28]

A Neural Network Architecture for Multilingual Punctuation Generation

Miguel Ballesteros and Leo Wanner

Even syntactically correct sentences are perceived as awkward if they do not contain correct punctuation. Still, the problem of automatic generation of punctuation marks has been largely neglected for a long time. We present a novel model introduces punctuation marks into raw text material using a transition-based stack LSTM model. Unlike the state-of-the-art approaches, our model is language-independent and also neutral with respect to the intended use of the punctuation. Multilingual experiments show that it achieves high accuracy on the full range of punctuation marks across languages.

[Summarization] [ID:S29]

Neural Headline Generation on Abstract Meaning Representation

Sho Takase, Jun Suzuki, Naoaki Okazaki, Tsutomu Hirao, and Masaaki Nagata

Neural network-based encoder-decoder models are among recent attractive methodologies for tackling natural language generation tasks. This paper investigates the usefulness of structural syntactic and semantic information additionally incorporated in a baseline neural attention-based model. We encode results obtained from an abstract meaning representation (AMR) parser using a modified version of Tree-LSTM. Our proposed attention-based AMR encoder-decoder model improves headline generation benchmarks compared with the baseline neural attention-based model.

[Text Mining & Applications] [ID:S30]

Robust Gram Embeddings

Taygun Kekec and David M. J. Tax

Word embedding models learn vectorial word representations that can be used in a variety of NLP applications. When training data is scarce, these models risk losing their generalization abilities due to the complexity of the models and the overfitting to finite data. We propose a regularized embedding formulation, called *Robust Gram* (RG), which penalizes overfitting by suppressing the disparity between target and context embeddings. Our experimental analysis shows that the RG model trained on small datasets generalizes better compared to alternatives, is more robust to variations in the training set, and correlates well to human similarities in a set of word similarity tasks.

[Text Mining & Applications] [ID:S31]

SimpleScience: Lexical Simplification of Scientific Terminology

Yea Seul Kim, Jessica Hullman, Matthew Burgess, and Eytan Adar

Lexical simplification of scientific terms represents a unique challenge due to the lack of a standard parallel corpora and fast rate at which vocabulary shift along with research. We introduce SimpleScience, a lexical simplification approach for scientific terminology. We use word embeddings to extract simplification rules from a parallel corpora containing scientific publications and Wikipedia. To evaluate our system we construct SimpleSciGold, a novel gold standard set for science-related simplifications. We find that our approach outperforms prior context-aware approaches at generating simplifications for scientific terms.

[Text Mining & Applications] [ID:S32]

Automatic Features for Essay Scoring — An Empirical Study

Fei Dong and Yue Zhang

Essay scoring is a complicated processing requiring analyzing, summarizing and judging expertise. Traditional work on essay scoring focused on automatic handcrafted features, which are expensive yet sparse. Neural models offer a way to learn syntactic and semantic features automatically, which can potentially improve upon discrete features. In this paper, we employ convolutional neural network (CNN) for the effect of automatically learning features, and compare the result with the state-of-art discrete baselines. For in-domain and domain-adaptation essay scoring tasks, our neural model empirically outperforms discrete models.

5

Main Conference: Thursday, November 3

Overview

07:30 – 17:30 **Registration Day 2** *Salon GHJ Prefunction*

08:00 – 09:00 **Morning Coffee** *Salon GHJ Prefunction*

09:00 – 10:00 **Session P2: Plenary Session II** *Salon FG*

09:00 – 10:00 *Invited Talk: Learning Models of Language, Action and Perception for Human-Robot Collaboration (Stefanie Tellex)*

10:00 – 10:30 **Coffee Break** *Salon GHJ Prefunction*

10:30 – 12:10 **Session 4**

Semantics and Semantic Parsing *Salon FG*	NLP for Social Science and Health *Salon J*	Language Models *Room 616 AB*

12:10 – 13:40 **Lunch**

13:00 – 13:40 **SIGDAT Business Meeting**

13:40 – 15:20 **Session 5**

Text Generation *Salon FG*	Discourse and Document Structure *Salon J*	Machine Translation and Multilingual Applications *Room 616 AB*

15:20 – 15:50 **Coffee Break** *Salon GHJ Prefunction*

15:50 – 17:30 **Session 6**

Neural Sequence-to-Sequence Models *Salon FG*	Text Mining and NLP Applications *Salon J*	Knowledge Base and Inference *Room 616 AB*

17:30 – 17:45 **Break**

17:45 – 18:15 **Poster Spotlight B (Half-minute Madness)** *Salon FG*

18:15 – 20:15 **Poster Session B** *Salon H and J*

Invited Speaker: Stefanie Tellex

Learning Models of Language, Action and Perception for Human-Robot Collaboration

Thursday, November 3, 2016,

Abstract: Robots can act as a force multiplier for people, whether a robot assisting an astronaut with a repair on the International Space station, a UAV taking flight over our cities, or an autonomous vehicle driving through our streets. To achieve complex tasks, it is essential for robots to move beyond merely interacting with people and toward collaboration, so that one person can easily and flexibly work with many autonomous robots. The aim of my research program is to create autonomous robots that collaborate with people to meet their needs by learning decision-theoretic models for communication, action, and perception. Communication for collaboration requires models of language that map between sentences and aspects of the external world. My work enables a robot to learn compositional models for word meanings that allow a robot to explicitly reason and communicate about its own uncertainty, increasing the speed and accuracy of human-robot communication. Action for collaboration requires models that match how people think and talk, because people communicate about all aspects of a robot's behavior, from low-level motion preferences (e.g., "Please fly up a few feet") to high-level requests (e.g., "Please inspect the building"). I am creating new methods for learning how to plan in very large, uncertain state-action spaces by using hierarchical abstraction. Perception for collaboration requires the robot to detect, localize, and manipulate the objects in its environment that are most important to its human collaborator. I am creating new methods for autonomously acquiring perceptual models in situ so the robot can perceive the objects most relevant to the human's goals. My unified decision-theoretic framework supports data-driven training and robust, feedback-driven human-robot collaboration.

Biography: Stefanie Tellex is an Assistant Professor of Computer Science and Assistant Professor of Engineering at Brown University. Her group, the Humans To Robots Lab, creates robots that seamlessly collaborate with people to meet their needs using language, gesture, and probabilistic inference, aiming to empower every person with a collaborative robot. She completed her Ph.D. at the MIT Media Lab in 2010, where she developed models for the meanings of spatial prepositions and motion verbs. Her postdoctoral work at MIT CSAIL focused on creating robots that understand natural language. She has published at SIGIR, HRI, RSS, AAAI, IROS, ICAPs and ICMI, winning Best Student Paper at SIGIR and ICMI, Best Paper at RSS, and an award from the CCC Blue Sky Ideas Initiative. Her awards include being named one of IEEE Spectrum's AI's 10 to Watch in 2013, the Richard B. Salomon Faculty Research Award at Brown University, a DARPA Young Faculty Award in 2015, and a 2016 Sloan Research Fellowship. Her work has been featured in the press on National Public Radio, MIT Technology Review, Wired UK and the Smithsonian. She was

named one of Wired UK's Women Who Changed Science In 2015 and listed as one of MIT Technology Review's Ten Breakthrough Technologies in 2016.

Session 4 Overview

	Track A	Track B	Track C
	Semantics and Semantic Parsing Salon FG	*NLP for Social Science and Health* Salon J	*Language Models* Room 616 AB
10:30	Semantic Parsing with Semi-Supervised Sequential Autoencoders *Kočiský, Melis, Grefenstette, Dyer, Ling, Blunsom, and Hermann*	Demographic Dialectal Variation in Social Media: A Case Study of African-American English *Blodgett, Green, and O'Connor*	Fast, Small and Exact: Infinite-order Language Modelling with Compressed Suffix Trees *Shareghi, Petri, Haffari, and Cohn*
10:55	Equation Parsing : Mapping Sentences to Grounded Equations *Roy, Upadhyay, and Roth*	Understanding Language Preference for Expression of Opinion and Sentiment: What do Hindi-English Speakers do on Twitter? *Rudra, Rijhwani, Begum, Bali, Choudhury, and Ganguly*	Convolutional Neural Network Language Models *Pham, Kruszewski, and Boleda*
11:20	Automatic Extraction of Implicit Interpretations from Modal Constructions *Sanders and Blanco*	Detecting and Characterizing Events *Chaney, Wallach, Connelly, and Blei*	Sparse Non-negative Matrix Language Modeling *Pelemans, Shazeer, and Chelba*
11:45	Understanding Negation in Positive Terms Using Syntactic Dependencies *Sarabi and Blanco*	Large-scale Analysis of Counseling Conversations: An Application of Natural Language Processing to Mental Health *Althoff, Clark, and Leskovec*	Generalizing and Hybridizing Count-based and Neural Language Models *Neubig and Dyer*

Session 4A: Semantics and Semantic Parsing

Salon FG

Semantic Parsing with Semi-Supervised Sequential Autoencoders

Tomáš Kočiský, Gábor Melis, Edward Grefenstette, Chris Dyer, Wang Ling, Phil Blunsom, and Karl Moritz Hermann 10:30–10:55

We present a novel semi-supervised approach for sequence transduction and apply it to semantic parsing. The unsupervised component is based on a generative model in which latent sentences generate the unpaired logical forms. We apply this method to a number of semantic parsing tasks focusing on domains with limited access to labelled training data and extend those datasets with synthetically generated logical forms.

Equation Parsing : Mapping Sentences to Grounded Equations

Subhro Roy, Shyam Upadhyay, and Dan Roth 10:55–11:20

Identifying mathematical relations expressed in text is essential to understanding a broad range of natural language text from election reports, to financial news, to sport commentaries to mathematical word problems. This paper focuses on identifying and understanding mathematical relations described within a single sentence. We introduce the problem of Equation Parsing – given a sentence, identify noun phrases which represent variables, and generate the mathematical equation expressing the relation described in the sentence. We introduce the notion of projective equation parsing and provide an efficient algorithm to parse text to projective equations. Our system makes use of a high precision lexicon of mathematical expressions and a pipeline of structured predictors, and generates correct equations in $70%$ of the cases. In $60%$ of the time, it also identifies the correct noun phrase –> variables mapping, significantly outperforming baselines. We also release a new annotated dataset for task evaluation.

Automatic Extraction of Implicit Interpretations from Modal Constructions

Jordan Sanders and Eduardo Blanco 11:20–11:45

This paper presents an approach to extract implicit interpretations from modal constructions. Importantly, our approach uses a deterministic procedure to normalize eventualities and generate potential interpretations. An annotation effort demonstrates that these interpretations are intuitive to humans and most modal constructions convey at least one interpretation. Experimental results show that the task is challenging but can be automated.

Understanding Negation in Positive Terms Using Syntactic Dependencies

Zahra Sarabi and Eduardo Blanco 11:45–12:10

This paper presents a two-step procedure to extract positive meaning from verbal negation. We first generate potential positive interpretations manipulating syntactic dependencies. Then, we score them according to their likelihood. Manual annotations show that positive interpretations are ubiquitous and intuitive to humans. Experimental results show that dependencies are better suited than semantic roles for this task, and automation is possible.

Session 4B: NLP for Social Science and Health

Salon J

Demographic Dialectal Variation in Social Media: A Case Study of African-American English
Su Lin Blodgett, Lisa Green, and Brendan O'Connor 10:30–10:55

Though dialectal language is increasingly abundant on social media, few resources exist for developing NLP tools to handle such language. We conduct a case study of dialectal language in online conversational text by investigating African-American English (AAE) on Twitter. We propose a distantly supervised model to identify AAE-like language from demographics associated with geo-located messages, and we verify that this language follows well-known AAE linguistic phenomena. In addition, we analyze the quality of existing language identification and dependency parsing tools on AAE-like text, demonstrating that they perform poorly on such text compared to text associated with white speakers. We also provide an ensemble classifier for language identification which eliminates this disparity and release a new corpus of tweets containing AAE-like language.

Understanding Language Preference for Expression of Opinion and Sentiment: What do Hindi-English Speakers do on Twitter?
Koustav Rudra, Shruti Rijhwani, Rafiya Begum, Kalika Bali, Monojit Choudhury, and Niloy Ganguly 10:55–11:20

Linguistic research on multilingual societies has indicated that there is usually a preferred language for expression of emotion and sentiment. Paucity of data has limited such studies to participant interviews and speech transcriptions from small groups of speakers. In this paper, we report a study on 430,000 unique tweets from Indian users, specifically Hindi-English bilinguals, to understand the language of preference, if any, for expressing opinion and sentiment. To this end, we develop classifiers for opinion detection in these languages, and further classifying opinionated tweets into positive, negative and neutral sentiments. Our study indicates that Hindi (i.e., the native language) is preferred over English for expression of negative opinion and swearing. As an aside, we explore some common pragmatic functions of code-switching through sentiment detection.

Detecting and Characterizing Events
Allison Chaney, Hanna Wallach, Matthew Connelly, and David Blei 11:20–11:45

Significant events are characterized by interactions between entities (such as countries, organizations, or individuals) that deviate from typical interaction patterns. Analysts, including historians, political scientists, and journalists, commonly read large quantities of text to construct an accurate picture of when and where an event happened, who was involved, and in what ways. In this paper, we present the Capsule model for analyzing documents to detect and characterize events of potential significance. Specifically, we develop a model based on topic modeling that distinguishes between topics that describe "business as usual" and topics that deviate from these patterns. To demonstrate this model, we analyze a corpus of over two million U.S. State Department cables from the 1970s. We provide an open-source implementation of an inference algorithm for the model and a pipeline for exploring its results.

Large-scale Analysis of Counseling Conversations: An Application of Natural Language Processing to Mental Health
Tim Althoff, Kevin Clark, and Jure Leskovec 11:45–12:10

Mental illness is one of the most pressing public health issues of our time. While counseling and psychotherapy can be effective treatments, our knowledge about how to conduct successful counseling conversations has been limited due to lack of large-scale data with labeled outcomes of the conversations. In this paper, we present a large-scale, quantitative study on the discourse of text-message-based counseling conversations. We develop a set of novel computational discourse analysis methods to measure how various linguistic aspects of conversations are correlated with conversation outcomes. Applying techniques such as sequence-based conversation models, language model comparisons, message clustering, and psycholinguistics-inspired word frequency analyses, we discover actionable conversation strategies that are associated with better conversation outcomes.

Session 4C: Language Models

Room 616 AB

Fast, Small and Exact: Infinite-order Language Modelling with Compressed Suffix Trees

Ehsan Shareghi, Matthias Petri, Gholamreza Haffari, and Trevor Cohn 10:30–10:55

Efficient methods for storing and querying are critical for scaling high-order m-gram language models to large corpora. We propose a language model based on compressed suffix trees, a representation that is highly compact and can be easily held in memory, while supporting queries needed in computing language model probabilities on-the-fly. We present several optimizations which improve query runtimes up to 2500x, despite only incurring a modest increase in construction time and memory usage. For large corpora and high Markov orders, our method is highly competitive with the state-of-the-art KenLM package. It imposes much lower memory requirements, often by orders of magnitude, and has runtimes that are either similar (for training) or comparable (for querying).

Convolutional Neural Network Language Models

Ngoc-Quan Pham, Germán Kruszewski, and Gemma Boleda 10:55–11:20

Convolutional Neural Networks (CNNs) have shown to yield very strong results in several Computer Vision tasks. Their application to language has received much less attention, and it has mainly focused on static classification tasks, such as sentence classification for Sentiment Analysis or relation extraction. In this work, we study the application of CNNs to language modeling, a dynamic, sequential prediction task that needs models to capture local as well as long-range dependency information. Our contribution is two fold. First, we show that CNNs achieve 13% higher absolute performance than feed-forward neural language models, demonstrating their potential for language representation even in sequential tasks. Second, we gain some understanding of the behavior of the model, showing that CNNs in language act as feature detectors at a high level of abstraction, as in Computer Vision, and that the model can profitably use information from as far as 16 words before the target.

Sparse Non-negative Matrix Language Modeling

Joris Pelemans, Noam Shazeer, and Ciprian Chelba 11:20–11:45

We present Sparse Non-negative Matrix (SNM) estimation, a novel probability estimation technique for language modeling that can efficiently incorporate arbitrary features. We evaluate SNM language models on two corpora: the One Billion Word Benchmark and a subset of the LDC English Gigaword corpus. Results show that SNM language models trained with n-gram features are a close match for the well-established Kneser-Ney models. The addition of skip-gram features yields a model that is in the same league as the state-of-the-art recurrent neural network language models, as well as complementary: combining the two modeling techniques yields the best known result on the One Billion Word Benchmark. On the Gigaword corpus further improvements are observed using features that cross sentence boundaries. The computational advantages of SNM estimation over both maximum entropy and neural network estimation are probably its main strength, promising an approach that has large flexibility in combining arbitrary features and yet scales gracefully to large amounts of data.

Generalizing and Hybridizing Count-based and Neural Language Models

Graham Neubig and Chris Dyer 11:45–12:10

Language models (LMs) are statistical models that calculate probabilities over sequences of words or other discrete symbols. Currently two major paradigms for language modeling exist: count-based n-gram models, which have advantages of scalability and test-time speed, and neural LMs, which often achieve superior modeling performance. We demonstrate how both varieties of models can be unified in a single modeling framework that defines a set of probability distributions over the vocabulary of words, and then dynamically calculates mixture weights over these distributions. This formulation allows us to create novel hybrid models that combine the desirable features of count-based and neural LMs, and experiments demonstrate the advantages of these approaches.

Session 5 Overview

	Track A	**Track B**	**Track C**
	Text Generation Salon FG	*Discourse and Document Structure* Salon J	*Machine Translation and Multilingual Applications* Room 616 AB
13:40	Reasoning about Pragmatics with Neural Listeners and Speakers *Andreas and Klein*	What makes a convincing argument? Empirical analysis and detecting attributes of convincingness in Web argumentation *Habernal and Gurevych*	An Unsupervised Probability Model for Speech-to-Translation Alignment of Low-Resource Languages *Anastasopoulos, Chiang, and Duong*
14:05	Generating Topical Poetry *Ghazvininejad, Shi, Choi, and Knight*	Recognizing Implicit Discourse Relations via Repeated Reading: Neural Networks with Multi-Level Attention *Liu and Li*	HUME: Human UCCA-Based Evaluation of Machine Translation *Birch, Abend, Bojar, and Haddow*
14:30	Deep Reinforcement Learning for Dialogue Generation *Li, Monroe, Ritter, Jurafsky, Galley, and Gao*	Antecedent Selection for Sluicing: Structure and Content *Anand and Hardt*	Improving Multilingual Named Entity Recognition with Wikipedia Entity Type Mapping *Ni and Florian*
14:55	Neural Text Generation from Structured Data with Application to the Biography Domain *Lebret, Grangier, and Auli*	Intra-Sentential Subject Zero Anaphora Resolution using Multi-Column Convolutional Neural Network *Iida, Torisawa, Oh, Kruengkrai, and Kloetzer*	Learning Crosslingual Word Embeddings without Bilingual Corpora *Duong, Kanayama, Ma, Bird, and Cohn*

Session 5A: Text Generation

Salon FG

Reasoning about Pragmatics with Neural Listeners and Speakers

Jacob Andreas and Dan Klein 13:40–14:05

We present a model for contrastively describing scenes, in which contrastive behavior results from a combination of inference-driven pragmatics and learned semantics. Like previous learned approaches to language generation, our model uses a simple feature-driven architecture (here a pair of neural "listener" and "speaker" models) to ground language in the world. Like inference-driven approaches to pragmatics, our model actively reasons about listener behavior when selecting utterances. For training, our approach requires only ordinary captions, annotated without demonstration of the pragmatic behavior the model ultimately exhibits. In human evaluations on a referring expression game, our approach succeeds 81% of the time, compared to 69% using existing techniques.

Generating Topical Poetry

Marjan Ghazvininejad, Xing Shi, Yejin Choi, and Kevin Knight 14:05–14:30

We describe Hafez, a program that generates any number of distinct poems on a user-supplied topic. Poems obey rhythmic and rhyme constraints. We describe the poetry-generation algorithm, give experimental data concerning its parameters, and show its generality with respect to language and poetic form.

Deep Reinforcement Learning for Dialogue Generation

Jiwei Li, Will Monroe, Alan Ritter, Dan Jurafsky, Michel Galley, and Jianfeng Gao 14:30–14:55

Recent neural models of dialogue generation offer great promise for generating responses for conversational agents, but tend to be shortsighted, predicting utterances one at a time while ignoring their influence on future outcomes. Modeling the future direction of a dialogue is crucial to generating coherent, interesting dialogues, a need which led traditional NLP models of dialogue to draw on reinforcement learning. In this paper, we show how to integrate these goals, applying deep reinforcement learning to model future reward in chatbot dialogue. The model simulates dialogues between two virtual agents, using policy gradient methods to reward sequences that display three useful conversational properties: informativity, coherence, and ease of answering (related to forward-looking function). We evaluate our model on diversity, length as well as with human judges, showing that the proposed algorithm generates more interactive responses and manages to foster a more sustained conversation in dialogue simulation. This work marks a first step towards learning a neural conversational model based on the long-term success of dialogues.

Neural Text Generation from Structured Data with Application to the Biography Domain

Rémi Lebret, David Grangier, and Michael Auli 14:55–15:20

This paper introduces a neural model for concept-to-text generation that scales to large, rich domains. It generates biographical sentences from fact tables on a new dataset of biographies from Wikipedia. This set is an order of magnitude larger than existing resources with over 700k samples and a 400k vocabulary. Our model builds on conditional neural language models for text generation. To deal with the large vocabulary, we extend these models to mix a fixed vocabulary with copy actions that transfer sample-specific words from the input database to the generated output sentence. To deal with structured data, we allow the model to embed words differently depending on the data fields in which they occur. Our neural model significantly outperforms a Templated Kneser-Ney language model by nearly 15 BLEU.

Session 5B: Discourse and Document Structure

Salon J

What makes a convincing argument? Empirical analysis and detecting attributes of convincingness in Web argumentation

Ivan Habernal and Iryna Gurevych 13:40–14:05

This article tackles a new challenging task in computational argumentation. Given a pair of two arguments to a certain controversial topic, we aim to directly assess qualitative properties of the arguments in order to explain why one argument is more convincing than the other one. We approach this task in a fully empirical manner by annotating 26k explanations written in natural language. These explanations describe convincingness of arguments in the given argument pair, such as their strengths or flaws. We create a new crowd-sourced corpus containing 9,111 argument pairs, multi-labeled with 17 classes, which was cleaned and curated by employing several strict quality measures. We propose two tasks on this data set, namely (1) predicting the full label distribution and (2) classifying types of flaws in less convincing arguments. Our experiments with feature-rich SVM learners and Bidirectional LSTM neural networks with convolution and attention mechanism reveal that such a novel fine-grained analysis of Web argument convincingness is a very challenging task. We release the new UKPConvArg2 corpus and software under permissive licenses to the research community.

Recognizing Implicit Discourse Relations via Repeated Reading: Neural Networks with Multi-Level Attention

Yang Liu and Sujian Li 14:05–14:30

Recognizing implicit discourse relations is a challenging but important task in the field of Natural Language Processing. For such a complex text processing task, different from previous studies, we argue that it is necessary to repeatedly read the arguments and dynamically exploit the efficient features useful for recognizing discourse relations. To mimic the repeated reading strategy, we propose the neural networks with multi-level attention (NNMA), combining the attention mechanism and external memories to gradually fix the attention on some specific words helpful to judging the discourse relations. Experiments on the PDTB dataset show that our proposed method achieves the state-of-art results. The visualization of the attention weights also illustrates the progress that our model observes the arguments on each level and progressively locates the important words.

Antecedent Selection for Sluicing: Structure and Content

Pranav Anand and Daniel Hardt 14:30–14:55

Sluicing is an elliptical process where the majority of a question can go unpronounced as long as there is a salient antecedent in previous discourse. This paper considers the task of antecedent selection: finding the correct antecedent for a given case of sluicing. We argue that both syntactic and discourse relationships are important in antecedent selection, and we construct linguistically sophisticated features that describe the relevant relationships. We also define features that describe the relation of the content of the antecedent and the sluice type. We develop a linear model which achieves accuracy of 72.4%, a substantial improvement over a strong manually constructed baseline. Feature analysis confirms that both syntactic and discourse features are important in antecedent selection.

Intra-Sentential Subject Zero Anaphora Resolution using Multi-Column Convolutional Neural Network

Ryu Iida, Kentaro Torisawa, Jong-Hoon Oh, Canasai Kruengkrai, and Julien Kloetzer 14:55–15:20

This paper proposes a method for intra-sentential subject zero anaphora resolution in Japanese. Our proposed method utilizes a Multi-column Convolutional Neural Network (MCNN) for predicting zero anaphoric relations. Motivated by Centering Theory and other previous works, we exploit as clues both the surface word sequence and the dependency tree of a target sentence in our MCNN. Even though the F-score of our method was lower than that of the state-of-the-art method, which achieved relatively high recall and low precision, our method achieved much higher precision (>0.8) in a wide range of recall levels. We believe such high precision is crucial for real-world NLP applications and thus our method is preferable to the state-of-the-art method.

Session 5C: Machine Translation and Multilingual Applications

Room 616 AB

An Unsupervised Probability Model for Speech-to-Translation Alignment of Low-Resource Languages

Antonios Anastasopoulos, David Chiang, and Long Duong 13:40–14:05

For many low-resource languages, spoken language resources are more likely to be annotated with translations than with transcriptions. Translated speech data is potentially valuable for documenting endangered languages or for training speech translation systems. A first step towards making use of such data would be to automatically align spoken words with their translations. We present a model that combines Dyer et al.'s reparameterization of IBM Model 2 (fast_align) and k-means clustering using Dynamic Time Warping as a distance metric. The two components are trained jointly using expectation-maximization. In an extremely low-resource scenario, our model performs significantly better than both a neural model and a strong baseline.

HUME: Human UCCA-Based Evaluation of Machine Translation

Alexandra Birch, Omri Abend, Ondřej Bojar, and Barry Haddow 14:05–14:30

Human evaluation of machine translation normally uses sentence-level measures such as relative ranking or adequacy scales. However, these provide no insight into possible errors, and do not scale well with sentence length. We argue for a semantics-based evaluation, which captures what meaning components are retained in the MT output, thus providing a more fine-grained analysis of translation quality, and enabling the construction and tuning of semantics-based MT. We present a novel human semantic evaluation measure, Human UCCA-based MT Evaluation (HUME), building on the UCCA semantic representation scheme. HUME covers a wider range of semantic phenomena than previous methods and does not rely on semantic annotation of the potentially garbled MT output. We experiment with four language pairs, demonstrating HUME's broad applicability, and report good inter-annotator agreement rates and correlation with human adequacy scores.

Improving Multilingual Named Entity Recognition with Wikipedia Entity Type Mapping

Jian Ni and Radu Florian 14:30–14:55

The state-of-the-art named entity recognition (NER) systems are statistical machine learning models that have strong generalization capability (i.e., can recognize unseen entities that do not appear in training data) based on lexical and contextual information. However, such a model could still make mistakes if its features favor a wrong entity type. In this paper, we utilize Wikipedia as an open knowledge base to improve multilingual NER systems. Central to our approach is the construction of high-accuracy, high-coverage multilingual Wikipedia entity type mappings. These mappings are built from weakly annotated data and can be extended to new languages with no human annotation or language-dependent knowledge involved. Based on these mappings, we develop several approaches to improve an NER system. We evaluate the performance of the approaches via experiments on NER systems trained for 6 languages. Experimental results show that the proposed approaches are effective in improving the accuracy of such systems on unseen entities, especially when a system is applied to a new domain or it is trained with little training data (up to 18.3 F_1 score improvement).

Learning Crosslingual Word Embeddings without Bilingual Corpora

Long Duong, Hiroshi Kanayama, Tengfei Ma, Steven Bird, and Trevor Cohn 14:55–15:20

Crosslingual word embeddings represent lexical items from different languages in the same vector space, enabling transfer of NLP tools. However, previous attempts had expensive resource requirements, difficulty incorporating monolingual data or were unable to handle polysemy. We address these drawbacks in our method which takes advantage of a high coverage dictionary in an EM style training algorithm over monolingual corpora in two languages. Our model achieves state-of-the-art performance on bilingual lexicon induction task exceeding models using large bilingual corpora, and competitive results on the monolingual word similarity and cross-lingual document classification task.

Session 6 Overview

	Track A	**Track B**	**Track C**
	Neural Sequence-to-Sequence Models Salon FG	*Text Mining and NLP Applications* Salon J	*Knowledge Base and Inference* Room 616 AB
15:50	Sequence-to-Sequence Learning as Beam-Search Optimization *Wiseman and Rush*	Poet Admits // Mute Cypher: Beam Search to find Mutually Enciphering Poetic Texts *Peterson and Fyshe*	AFET: Automatic Fine-Grained Entity Typing by Hierarchical Partial-Label Embedding *Ren, He, Qu, Huang, Ji, and Han*
16:15	Online Segment to Segment Neural Transduction *Yu, Buys, and Blunsom*	All Fingers are not Equal: Intensity of References in Scientific Articles *Chakraborty and Narayanam*	Mining Inference Formulas by Goal-Directed Random Walks *Wei, Zhao, and Liu*
16:40	Sequence-Level Knowledge Distillation *Kim and Rush*	Improving Users' Demographic Prediction via the Videos They Talk about *Wang, Xiao, Ma, and Xiao*	Lifted Rule Injection for Relation Embeddings *Demeester, Rocktäschel, and Riedel*
17:05	Controlling Output Length in Neural Encoder-Decoders *Kikuchi, Neubig, Sasano, Takamura, and Okumura*	Understanding Satirical Articles Using Common-Sense *Goldwasser and Zhang*	Key-Value Memory Networks for Directly Reading Documents *Miller, Fisch, Dodge, Karimi, Bordes, and Weston*

Session 6A: Neural Sequence-to-Sequence Models

Salon FG

Sequence-to-Sequence Learning as Beam-Search Optimization

Sam Wiseman and Alexander M. Rush 15:50–16:15

Sequence-to-Sequence (seq2seq) modeling has rapidly become an important general-purpose NLP tool that has proven effective for many text-generation and sequence-labeling tasks. Seq2seq builds on deep neural language modeling and inherits its remarkable accuracy in estimating local, next-word distributions. In this work, we introduce a model and beam-search training scheme, based on the work of Daume III and Marcu (2005), that extends seq2seq to learn global sequence scores. This structured approach avoids classical biases associated with local training and unifies the training loss with the test-time usage, while preserving the proven model architecture of seq2seq and its efficient training approach. We show that our system outperforms a highly-optimized attention-based seq2seq system and other baselines on three different sequence to sequence tasks: word ordering, parsing, and machine translation.

Online Segment to Segment Neural Transduction

Lei Yu, Jan Buys, and Phil Blunsom 16:15–16:40

We introduce an online neural sequence to sequence model that learns to alternate between encoding and decoding segments of the input as it is read. By independently tracking the encoding and decoding representations our algorithm permits exact polynomial marginalization of the latent segmentation during training, and during decoding beam search is employed to find the best alignment path together with the predicted output sequence. Our model tackles the bottleneck of vanilla encoder-decoders that have to read and memorize the entire input sequence in their fixed-length hidden states before producing any output. It is different from the previous attentive models in that, instead of treating the attention weights as output of a deterministic function, our model assigns attention weights to a sequential latent variable which can be marginalized out and permits online generation. Experiments on abstractive sentence summarization and morphological inflection show significant performance gains over the baseline encoder-decoders.

Sequence-Level Knowledge Distillation

Yoon Kim and Alexander M. Rush 16:40–17:05

Neural machine translation (NMT) offers a novel alternative formulation of translation that is potentially simpler than statistical approaches. However to reach competitive performance, NMT models need to be exceedingly large. In this paper we consider applying knowledge distillation approaches (Bucila et al., 2006, Hinton et al., 2015) that have proven successful for reducing the size of neural models in other domains to the problem of NMT. We demonstrate that standard knowledge distillation applied to word-level prediction can be effective for NMT, and also introduce two novel sequence-level versions of knowledge distillation that further improve performance, and somewhat surprisingly, seem to eliminate the need for beam search (even when applied on the original teacher model). Our best student model runs 10 times faster than its state-of-the-art teacher with little loss in performance. It is also significantly better than a baseline model trained without knowledge distillation: by 4.2/1.7 BLEU with greedy decoding/beam search. Applying weight pruning on top of knowledge distillation results in a student model that has 13 times fewer parameters than the original teacher model, with a decrease of 0.4 BLEU.

Controlling Output Length in Neural Encoder-Decoders

Yuta Kikuchi, Graham Neubig, Ryohei Sasano, Hiroya Takamura, and Manabu Okumura 17:05–17:30

Neural encoder-decoder models have shown great success in many sequence generation tasks. However, previous work has not investigated situations in which we would like control the length of encoder-decoder outputs. This capability is crucial for applications such as text summarization, in which we have to generate concise summaries with a desired length. In this paper, we propose methods for controlling the output sequence length for neural encoder-decoder models: two decoding-based methods and two learning-based methods. Results show that our learning-based methods have the capability to control length without degrading summary quality in a summarization task.

Session 6B: Text Mining and NLP Applications

Salon J

Poet Admits // Mute Cypher: Beam Search to find Mutually Enciphering Poetic Texts

Cole Peterson and Alona Fyshe 15:50–16:15

The Xenotext Experiment implants poetry into an extremophile's DNA, and uses that DNA to generate new poetry in a protein form. The molecular machinery of life requires that these two poems encipher each other under a symmetric substitution cipher. We search for ciphers which permit writing under the Xenotext constraints, incorporating ideas from cipher-cracking algorithms, and using n-gram data to assess a cipher's "writability". Our algorithm, Beam Verse, is a beam search which uses new heuristics to navigate the cipher-space. We find thousands of ciphers which score higher than successful ciphers used to write Xenotext constrained texts.

All Fingers are not Equal: Intensity of References in Scientific Articles

Tanmoy Chakraborty and Ramasuri Narayanam 16:15–16:40

Research accomplishment is usually measured by considering all citations with equal importance, thus ignoring the wide variety of purposes an article is being cited for. Here, we posit that measuring the intensity of a reference is crucial not only to perceive better understanding of research endeavor, but also to improve the quality of citation-based applications. To this end, we collect a rich annotated dataset with references labeled by the intensity, and propose a novel graph-based semi-supervised model, GraLap to label the in- tensity of references. Experiments with AAN datasets show a significant improvement compared to the baselines to achieve the true labels of the references (46% better correlation). Finally, we provide four applications to demonstrate how the knowledge of reference intensity leads to design better real-world applications.

Improving Users' Demographic Prediction via the Videos They Talk about

Yuan Wang, Yang Xiao, Chao Ma, and Zhen Xiao 16:40–17:05

In this paper, we improve microblog users' demographic prediction by fully utilizing their video related behaviors. First, we collect the describing words of currently popular videos, including video names, actor names and video keywords, from video websites. Secondly, we search these describing words in users' microblogs, and build the direct relationships between users and the appeared words. After that, to make the sparse relationship denser, we propose a Bayesian method to calculate the probability of connections between users and other video describing words. Lastly, we build two models to predict users' demographics with the obtained direct and indirect relationships. Based on a large real-world dataset, experiment results show that our method can significantly improve these words' demographic predictive ability.

Understanding Satirical Articles Using Common-Sense

Dan Goldwasser and Xiao Zhang 17:05–17:30

Automatic satire detection is a subtle text classification task, for machines and at times, even for humans. In this paper we argue that satire detection should be approached using common-sense inferences, rather than traditional text classification methods. We present a highly structured latent variable model capturing the required inferences. The model abstracts over the specific entities appearing in the articles, grouping them into generalized categories, thus allowing the model to adapt to previously unseen situations.

Session 6C: Knowledge Base and Inference

Room 616 AB

AFET: Automatic Fine-Grained Entity Typing by Hierarchical Partial-Label Embedding

Xiang Ren, Wenqi He, Meng Qu, Lifu Huang, Heng Ji, and Jiawei Han 15:50–16:15

Distant supervision has been widely used in current systems of fine-grained entity typing to automatically assign categories (entity types) to entity mentions. However, the types so obtained from knowledge bases are often incorrect for the entity mention's local context. This paper proposes a novel embedding method to separately model "clean" and "noisy" mentions, and incorporates the given type hierarchy to induce loss functions. We formulate a joint optimization problem to learn embeddings for mentions and type-paths, and develop an iterative algorithm to solve the problem. Experiments on three public datasets demonstrate the effectiveness and robustness of the proposed method, with an average 15% improvement in accuracy over the next best compared method.

Mining Inference Formulas by Goal-Directed Random Walks

Zhuoyu Wei, Jun Zhao, and Kang Liu 16:15–16:40

Deep inference on a large-scale knowledge base (KB) need a mass of formulas, but it is almost impossible to create all formulas manually. Data-driven methods have been proposed to mine formulas from KBs automatically, where random sampling and approximate calculation are common techniques to handle big data. Thereinto, Random Walk is the most suitable for knowledge graph data. However, a pure random walk without goals has a poor efficiency of mining useful formulas, and even introduces lots of noise which may mislead inference. Although several heuristic rules have been proposed to direct random walks, they do not work well due to the diversity of formulas. To this end, we propose a novel goal-directed algorithm, which directs random walks by the specific inference target at each step. The algorithm is more inclined to visit beneficial structures to infer the target, so it can increase efficiency of random walks and avoid noise simultaneously. In our experiments, we prove our approach is high-efficiency and perform best on the KB link prediction task.

Lifted Rule Injection for Relation Embeddings

Thomas Demeester, Tim Rocktäschel, and Sebastian Riedel 16:40–17:05

Methods based on representation learning currently hold the state-of-the-art in many natural language processing and knowledge base inference tasks. Yet, a major challenge is how to efficiently incorporate commonsense knowledge into such models. A recent approach regularizes relation and entity representations by propositionalization of first-order logic rules. However, propositionalization does not scale beyond domains with only few entities and rules. In this paper we present a highly efficient method for incorporating implication rules into distributed representations for automated knowledge base construction. We map entity-tuple embeddings into an approximately Boolean space and encourage a partial ordering over relation embeddings based on implication rules mined from WordNet. Surprisingly, we find that the strong restriction of the entity-tuple embedding space does not hurt the expressiveness of the model and even acts as a regularizer that improves generalization. By incorporating few commonsense rules, we achieve an increase of 2 percentage points mean average precision over a matrix factorization baseline, while observing a negligible increase in runtime.

Key-Value Memory Networks for Directly Reading Documents

Alexander Miller, Adam Fisch, Jesse Dodge, Amir-Hossein Karimi, Antoine Bordes, and Jason Weston 17:05–17:30

Directly reading documents and being able to answer questions from them is an unsolved challenge. To avoid its inherent difficulty, question answering (QA) has been directed towards using Knowledge Bases (KBs) instead, which has proven effective. Unfortunately KBs often suffer from being too restrictive, as the schema cannot support certain types of answers, and too sparse, e.g. Wikipedia contains much more information than Freebase. In this work we introduce a new method, Key-Value Memory Networks, that makes reading documents more viable by utilizing different encodings in the addressing and output stages of the memory read operation. To compare using KBs, information extraction or Wikipedia documents directly in a single framework we construct an analysis tool, WikiMovies, a QA dataset that contains raw text alongside a preprocessed KB, in the domain of movies. Our method reduces the gap between all three settings. It also achieves state-of-the-art results on the existing WikiQA benchmark.

Poster Spotlight B (Half-minute Madness)

Chair: Brendan O'Connor Joel Tetreault

Salon H and J *17:45–18:15*

Poster Session B

Salon H and J *18:15–20:15*

[Discourse & Dialogue] [ID:L01]

Analyzing Framing through the Casts of Characters in the News

Dallas Card, Justin Gross, Amber Boydstun, and Noah A. Smith

We present an unsupervised model for the discovery and clustering of latent "personas" (characterizations of entities). Our model simultaneously clusters documents featuring similar collections of personas. We evaluate this model on a collection of news articles about immigration, showing that personas help predict the coarse-grained framing annotations in the Media Frames Corpus. We also introduce automated model selection as a fair and robust form of evaluation.

[Discourse & Dialogue] [ID:L02]

The Teams Corpus and Entrainment in Multi-Party Spoken Dialogues

Diane Litman, Susannah Paletz, Zahra Rahimi, Stefani Allegretti, and Caitlin Rice

When interacting individuals entrain, they begin to speak more like each other. To support research on entrainment in cooperative multi-party dialogues, we have created a corpus where teams of three or four speakers play two rounds of a cooperative board game. We describe the experimental design and technical infrastructure used to collect our corpus, which consists of audio, video, transcriptions, and questionnaire data for 63 teams (47 hours of audio). We illustrate the use of our corpus as a novel resource for studying team entrainment by 1) developing and evaluating team-level acoustic-prosodic entrainment measures that extend existing dyad measures, and 2) investigating relationships between team entrainment and participation dominance.

[Discourse & Dialogue] [ID:L03]

Personalized Emphasis Framing for Persuasive Message Generation

Tao Ding and Shimei Pan

In this paper, we present a study on personalized emphasis framing which can be used to tailor the content of a message to enhance its appeal to different individuals. With this framework, we directly model content selection decisions based on a set of psychologically-motivated domain-independent personal traits including personality (e.g., extraversion and conscientiousness) and basic human values (e.g., self-transcendence and hedonism). We also demonstrate how the analysis results can be used in automated personalized content selection for persuasive message generation.

[Information Extraction] [ID:L04]

Cross Sentence Inference for Process Knowledge

Samuel Louvan, Chetan Naik, Sadhana Kumaravel, Heeyoung Kwon, Niranjan Balasubramanian, and Peter Clark

For AI systems to reason about real world situations, they need to recognize which processes are at play and which entities play key roles in them. Our goal is to extract this kind of role-based knowledge about processes, from multiple sentence-level descriptions. This knowledge is hard to acquire; while semantic role labeling (SRL) systems can extract sentence level role information about individual mentions of a

process, their results are often noisy and they do not attempt create a globally consistent characterization of a process. To overcome this, we extend standard within sentence joint inference to inference across multiple sentences. This cross sentence inference promotes role assignments that are compatible across different descriptions of the same process. When formulated as an Integer Linear Program, this leads to improvements over within-sentence inference by nearly 3 % in F1. The resulting role-based knowledge is of high quality (with a F1 of nearly 82).

[Information Extraction] [ID:L05]

Toward Socially-Infused Information Extraction: Embedding Authors, Mentions, and Entities

Yi Yang, Ming-Wei Chang, and Jacob Eisenstein

Entity linking is the task of identifying mentions of entities in text, and linking them to entries in a knowledge base. This task is especially difficult in microblogs, as there is little additional text to provide disambiguating context; rather, authors rely on an implicit common ground of shared knowledge with their readers. In this paper, we attempt to capture some of this implicit context by exploiting the social network structure in microblogs. We build on the theory of homophily, which implies that socially linked individuals share interests, and are therefore likely to mention the same sorts of entities. We implement this idea by encoding authors, mentions, and entities in a continuous vector space, which is constructed so that socially-connected authors have similar vector representations. These vectors are incorporated into a neural structured prediction model, which captures structural constraints that are inherent in the entity linking task. Together, these design decisions yield F1 improvements of 1%-5% on benchmark datasets, as compared to the previous state-of-the-art.

[Information Extraction] [ID:L06]

Phonologically Aware Neural Model for Named Entity Recognition in Low Resource Transfer Settings

Akash Bharadwaj, David Mortensen, Chris Dyer, and Jaime Carbonell

Named Entity Recognition is a well established information extraction task with many state of the art systems existing for a variety of languages. Most systems rely on language specific resources, large annotated corpora, gazetteers and feature engineering to perform well monolingually. In this paper, we introduce an attentional neural model which only uses language universal phonological character representations with word embeddings to achieve state of the art performance in a monolingual setting using supervision and which can quickly adapt to a new language with minimal or no data. We demonstrate that phonological character representations facilitate cross-lingual transfer, outperform orthographic representations and incorporating both attention and phonological features improves statistical efficiency of the model in 0-shot and low data transfer settings with no task specific feature engineering in the source or target language.

[Language Modeling] [ID:L07]

Long-Short Range Context Neural Networks for Language Modeling

Youssef Oualil, Mittul Singh, Clayton Greenberg, and Dietrich Klakow

The goal of language modeling techniques is to capture the statistical and structural properties of natural languages from training corpora. This task typically involves the learning of short range dependencies, which generally model the syntactic properties of a language and/or long range dependencies, which are semantic in nature. We propose in this paper a new multi-span architecture, which separately models the short and long context information while it dynamically merges them to perform the language modeling task. This is done through a novel recurrent Long-Short Range Context (LSRC) network, which explicitly models the local (short) and global (long) context using two separate hidden states that evolve in time. This new architecture is an adaptation of the Long-Short Term Memory network (LSTM) to take into account the linguistic properties. Extensive experiments conducted on the Penn Treebank (PTB) and the Large Text Compression Benchmark (LTCB) corpus showed a significant reduction of the perplexity when compared to state-of-the-art language modeling techniques.

[Language and Vision] [ID:L08]

Jointly Learning Grounded Task Structures from Language Instruction and Visual Demonstration

Changsong Liu, Shaohua Yang, Sari Saba-Sadiya, Nishant Shukla, Yunzhong He, Song-chun Zhu, and Joyce Chai

To enable language-based communication and collaboration with cognitive robots, this paper presents an approach where an agent can learn task models jointly from language instruction and visual demonstration using an And-Or Graph (AoG) representation. The learned AoG captures a hierarchical task structure where linguistic labels (for language communication) are grounded to corresponding state changes

from the physical environment (for perception and action). Our empirical results on a cloth-folding domain have shown that, although state detection through visual processing is full of uncertainties and error prone, by a tight integration with language the agent is able to learn an effective AoG for task representation. The learned AoG can be further applied to infer and interpret on-going actions from new visual demonstration using linguistic labels at different levels of granularity.

[Language and Vision] [ID:L09]

Resolving Language and Vision Ambiguities Together: Joint Segmentation & Prepositional Attachment Resolution in Captioned Scenes

Gordon Christie, Ankit Laddha, Aishwarya Agrawal, Stanislaw Antol, Yash Goyal, Kevin Kochersberger, and Dhruv Batra

We present an approach to simultaneously perform semantic segmentation and prepositional phrase attachment resolution for captioned images. The motivation for this work comes from the fact that some ambiguities in language simply cannot be resolved without simultaneously reasoning about an associated image. If we consider the sentence "I shot an elephant in my pajamas", looking at the language alone (and not reasoning about common sense), it is unclear if it is the person or the elephant that is wearing the pajamas or both. Our approach involves producing a diverse set of plausible hypotheses for both semantic segmentation and prepositional phrase attachment resolution that are then jointly reranked to select the most consistent pair. We show that our semantic segmentation and prepositional phrase attachment resolution modules have complementary strengths, and that joint reasoning produces more accurate results than any module operating in isolation. We also show that multiple hypotheses are crucial to improved multiple-module reasoning. Our vision and language approach significantly outperforms the Stanford Parser by 17.91% (28.69% relative) in one experiment, and by 12.83% (25.28% relative) in another. We also make small improvements over a vision system (DeepLab-CRF).

[Machine Learning] [ID:L10]

Charagram: Embedding Words and Sentences via Character n-grams

John Wieting, Mohit Bansal, Kevin Gimpel, and Karen Livescu

We present Charagram embeddings, a simple approach for learning character-based compositional models to embed textual sequences. A word or sentence is represented using a character n-gram count vector, followed by a single nonlinear transformation to yield a low-dimensional embedding. We use three tasks for evaluation: word similarity, sentence similarity, and part-of-speech tagging. We demonstrate that Charagram embeddings outperform more complex architectures based on character-level recurrent and convolutional neural networks, achieving new state-of-the-art performance on several similarity tasks.

[Machine Learning] [ID:L11]

Length bias in Encoder Decoder Models and a Case for Global Conditioning

Pavel Sountsov and Sunita Sarawagi

Encoder-decoder networks are popular for modeling sequences probabilistically in many applications. These models use the power of the Long Short-Term Memory (LSTM) architecture to capture the full dependence among variables, unlike earlier models like CRFs that typically assumed conditional independence among non-adjacent variables. However in practice encoder-decoder models exhibit a bias towards short sequences that surprisingly gets worse with increasing beam size. In this paper we show that such phenomenon is due to a discrepancy between the full sequence margin and the per-element margin enforced by the locally conditioned training objective of a encoder-decoder model. The discrepancy more adversely impacts long sequences, explaining the bias towards predicting short sequences. For the case where the predicted sequences come from a closed set, we show that a globally conditioned model alleviates the above problems of encoder-decoder models. From a practical point of view, our proposed model also eliminates the need for a beam-search during inference, which reduces to an efficient dot-product based search in a vector-space.

[Machine Learning] [ID:L12]

Comparing Apples to Apple: The Effects of Stemmers on Topic Models

Alexandra Schofield and David Mimno

Rule-based stemmers such as the Porter stemmer are frequently used to preprocess English corpora for topic modeling. In this work, we train and evaluate topic models on a variety of corpora using several different stemming algorithms. We examine several different quantitative measures of the resulting models, including likelihood, coherence, model stability, and entropy. Despite their frequent use in topic modeling, we find that stemmers produce no meaningful improvement in likelihood and coherence and in fact can degrade topic stability.

[Machine Translation] [ID:L13]

Does String-Based Neural MT Learn Source Syntax?
Xing Shi, Inkit Padhi, and Kevin Knight

We investigate whether a neural, encoder-decoder translation system learns syntactic information on the source side as a by-product of training. We propose two methods to detect whether the encoder has learned local and global source syntax. A fine-grained analysis of the syntactic structure learned by the encoder reveals which kinds of syntax are learned and which are missing.

[Machine Translation] [ID:L14]
Exploiting Source-side Monolingual Data in Neural Machine Translation
Jiajun Zhang and Chengqing Zong

Neural Machine Translation (NMT) based on the encoder-decoder architecture has recently become a new paradigm. Researchers have proven that the target-side monolingual data can greatly enhance the decoder model of NMT. However, the source-side monolingual data is not fully explored although it should be useful to strengthen the encoder model of NMT, especially when the parallel corpus is far from sufficient. In this paper, we propose two approaches to make full use of the source-side monolingual data in NMT. The first approach employs the self-learning algorithm to generate the synthetic large-scale parallel data for NMT training. The second approach applies the multi-task learning framework using two NMTs to predict the translation and the reordered source-side monolingual sentences simultaneously. The extensive experiments demonstrate that the proposed methods obtain significant improvements over the strong attention-based NMT.

[Machine Translation] [ID:L15]
Phrase-based Machine Translation is State-of-the-Art for Automatic Grammatical Error Correction
Marcin Junczys-Dowmunt and Roman Grundkiewicz

In this work, we study parameter tuning towards the M2 metric, the standard metric for automatic grammar error correction (GEC) tasks. After implementing M2 as a scorer in the Moses tuning framework, we investigate interactions of dense and sparse features, different optimizers, and tuning strategies for the CoNLL-2014 shared task. We notice erratic behavior when optimizing sparse feature weights with M2 and offer partial solutions. We find that a bare-bones phrase-based SMT setup with task-specific parameter-tuning outperforms all previously published results for the CoNLL-2014 test set by a large margin (46.37% M2 over previously 41.75%, an SMT system with neural features) while being trained on the same, publicly available data. Our newly introduced dense and sparse features widen that gap, and we improve the state-of-the-art to 49.49% M2.

[Machine Translation] [ID:L16]
Incorporating Discrete Translation Lexicons into Neural Machine Translation
Philip Arthur, Graham Neubig, and Satoshi Nakamura

Neural machine translation (NMT) often makes mistakes in translating low-frequency content words that are essential to understandng the meaning of the sentence. We propose a method to alleviate this problem by augmenting NMT systems with discrete translation lexicons that efficiently encode translations of these low-frequency words. We describe a method to calculate the lexicon probability of the next word in the translation candidate by using the attention vector of the NMT model to select which source word lexical probabilities the model should focus on. We test two methods to combine this probability with the standard NMT probability: (1) using it as a bias, and (2) linear interpolation. Experiments on two corpora show an improvement of 2.0-2.3 BLEU and 0.13-0.44 NIST score, and faster convergence time.

[Machine Translation] [ID:L17]
Transfer Learning for Low-Resource Neural Machine Translation
Barret Zoph, Deniz Yuret, Jonathan May, and Kevin Knight

The encoder-decoder framework for neural machine translation (NMT) has been shown effective in large data scenarios, but is much less effective for low-resource languages. We present a transfer learning method that significantly improves BLEUscores across a range of low-resource languages. Our key idea is to first train a high-resource language pair (the parent model), then transfer some of the learned parameters to the low-resource pair (the child model) to initialize and constrain training. Using our transfer learning method we improve baseline NMT models by an average of 5.6 BLEU on four low-resource language pairs. Ensembling and unknown word replacement add another 2 BLEU which brings the NMT performance on low-resource machine translation close to a strong syntax based machine translation (SBMT) system, exceeding its performance on one language pair. Additionally, using the transfer learning model for re-scoring, we can improve the SBMT system by an average of 1.3 BLEU, improving the state-of-the-art on low-resource machine translation.

[Question Answering] [ID:L18]

MixKMeans: Clustering Question-Answer Archives

Deepak P

Community-driven Question Answering (CQA) systems that crowdsource experiential information in the form of questions and answers and have accumulated valuable reusable knowledge. Clustering of QA datasets from CQA systems provides a means of organizing the content to ease tasks such as manual curation and tagging. In this paper, we present a clustering method that exploits the two-part question-answer structure in QA datasets to improve clustering quality. Our method, *MixKMeans*, composes question and answer space similarities in a way that the space on which the match is higher is allowed to dominate. This construction is motivated by our observation that semantic similarity between question-answer data (QAs) could get localized in either space. We empirically evaluate our method on a variety of real-world labeled datasets. Our results indicate that our method significantly outperforms state-of-the-art clustering methods for the task of clustering question-answer archives.

[Question Answering] [ID:L19]

It Takes Three to Tango: Triangulation Approach to Answer Ranking in Community Question Answering

Preslav Nakov, Lluís Màrquez, and Francisco Guzmán

We address the problem of answering new questions in community forums, by selecting suitable answers to already asked questions. We approach the task as an answer ranking problem, adopting a pairwise neural network architecture that selects which of two competing answers is better. We focus on the utility of the three types of similarities occurring in the triangle formed by the original question, the related question, and an answer to the related comment, which we call relevance, relatedness, and appropriateness. Our proposed neural network models the interactions among all input components using syntactic and semantic embeddings, lexical matching, and domain-specific features. It achieves state-of-the-art results, showing that the three similarities are important and need to be modeled together. Our experiments demonstrate that all feature types are relevant, but the most important ones are the lexical similarity features, the domain-specific features, and the syntactic and semantic embeddings.

[Question Answering] [ID:L20]

Character-Level Question Answering with Attention

Xiaodong He and David Golub

We show that a character-level encoder-decoder framework can be successfully applied to question answering with a structured knowledge base. We use our model for single-relation question answering and demonstrate the effectiveness of our approach on the SimpleQuestions dataset (Bordes et al., 2015), where we improve state-of-the-art accuracy from 63.9% to 70.9%, without use of ensembles. Importantly, our character-level model has 16x fewer parameters than an equivalent word-level model, can be learned with significantly less data compared to previous work, which relies on data augmentation, and is robust to new entities in testing.

[Question Answering] [ID:L21]

Learning to Generate Textual Data

Guillaume Bouchard, Pontus Stenetorp, and Sebastian Riedel

To learn text understanding models requiring millions of parameters, one needs a massive amount of data. Instead, we argue that generating data can compensate the need for large datasets. While defining generic data generators is tricky, we propose to allow these generators to be "weakly" specified, leaving letting the undetermined coefficients to be learned from data. We derived an efficient algorithm called GeneRe, that jointly estimate the parameters of the model and the undetermined sampling coefficients, removing the need for costly cross-validation. We illustrate its benefit by learning to solve math exam questions using a sequence-to-sequence recurrent network.

[Question Answering] [ID:L22]

A Theme-Rewriting Approach for Generating Algebra Word Problems

Rik Koncel-Kedziorski, Ioannis Konstas, Luke Zettlemoyer, and Hannaneh Hajishirzi

Texts present coherent stories that have a particular theme or overall setting, for example science fiction or westerns. In this paper, we present a text generation method called 'rewriting' that edits existing human-authored narratives to change their theme without changing the underlying story. We apply the approach to math word problems, where it might help students stay more engaged by quickly transforming all of their homework assignments to the theme of their favorite movie without changing the math concepts that are being taught. Our rewriting method uses a two-stage decoding process, which proposes new words from the target theme and scores the resulting stories according to a number of factors defining aspects of syntactic, semantic, and thematic coherence. Experiments demonstrate that

the final stories typically represent the new theme well while still testing the original math concepts, outperforming a number of baselines. We also release a new dataset of human-authored rewrites of math word problems in several themes.

[Sentiment Analysis] [ID:L23]

Context-Sensitive Lexicon Features for Neural Sentiment Analysis

Zhiyang Teng, Duy Tin Vo, and Yue Zhang

Sentiment lexicons have been leveraged as a useful source of features for sentiment analysis models, leading to the state-of-the-art accuracies. On the other hand, most existing methods use sentiment lexicons without considering context, typically taking the count, sum of strength, or maximum sentiment scores over the whole input. We propose a context-sensitive lexicon-based method based on a simple weighted-sum model, using a recurrent neural network to learn the sentiments strength, intensification and negation of lexicon sentiments in composing the sentiment value of sentences. Results show that our model can not only learn such operation details, but also give significant improvements over state-of-the-art recurrent neural network baselines without lexical features, achieving the best results on a Twitter benchmark.

[Sentiment Analysis] [ID:L24]

Event-Driven Emotion Cause Extraction with Corpus Construction

Lin Gui, Dongyin Wu, Ruifeng Xu, Qin Lu, and Yu Zhou

In this paper, we present our work in emotion cause extraction. Since there is no open dataset available, the lack of annotated resources has limited the research in this area. Thus, we first present a dataset we built using SINA city news. The annotation is based on the scheme of the W3C Emotion Markup Language. Second, we propose a 7-tuple definition to describe emotion cause events. Based on this general definition, we propose a new event-driven emotion cause extraction method using multi-kernel SVMs where a syntactical tree based approach is used to represent events in text. A convolution kernel based multi-kernel SVM are used to extract emotion causes. Because traditional convolution kernels do not use lexical information at the terminal nodes of syntactic trees, we modify the kernel function with a synonym based improvement. Even with very limited training data, we can still extract sufficient features for the task. Evaluations show that our approach achieves 11.6% higher F-measure compared to referenced methods. The contributions of our work include resource construction, concept definition and algorithm development.

[Sentiment Analysis] [ID:L25]

Neural Sentiment Classification with User and Product Attention

Huimin Chen, Maosong Sun, Cunchao Tu, Yankai Lin, and Zhiyuan Liu

Document-level sentiment classification aims to predict user's overall sentiment in a document about a product. However, most of existing methods only focus on local text information and ignore the global user preference and product characteristics. Even though some works take such information into account, they usually suffer from high model complexity and only consider word-level preference rather than semantic levels. To address this issue, we propose a hierarchical neural network to incorporate global user and product information into sentiment classification. Our model first builds a hierarchical LSTM model to generate sentence and document representations. Afterwards, user and product information is considered via attentions over different semantic levels due to its ability of capturing crucial semantic components. The experimental results show that our model achieves significant and consistent improvements compared to all state-of-the-art methods. The source code of this paper can be obtained from https://github.com/thunlp/NSC.

[Sentiment Analysis] [ID:L26]

Cached Long Short-Term Memory Neural Networks for Document-Level Sentiment Classification

Jiacheng Xu, Danlu Chen, Xipeng Qiu, and Xuanjing Huang

Recently, neural networks have achieved great success on sentiment classification due to their ability to alleviate feature engineering. However, one of the remaining challenges is to model long texts in document-level sentiment classification under a recurrent architecture because of the deficiency of the memory unit. To address this problem, we present a Cached Long Short-Term Memory neural networks (CLSTM) to capture the overall semantic information in long texts. CLSTM introduces a cache mechanism, which divides memory into several groups with different forgetting rates and thus enables the network to keep sentiment information better within a recurrent unit. The proposed CLSTM outperforms the state-of-the-art models on three publicly available document-level sentiment analysis datasets.

[Sentiment Analysis] [ID:L27]

Deep Neural Networks with Massive Learned Knowledge
Zhiting Hu, Zichao Yang, Ruslan Salakhutdinov, and Eric Xing

Regulating deep neural networks (DNNs) with human structured knowledge has shown to be of great benefit for improved accuracy and interpretability. We develop a general framework that enables learning knowledge and its confidence jointly with the DNNs, so that the vast amount of fuzzy knowledge can be incorporated and automatically optimized with little manual efforts. We apply the framework to sentence sentiment analysis, augmenting a DNN with massive linguistic constraints on discourse and polarity structures. Our model substantially enhances the performance using less training data, and shows improved interpretability. The principled framework can also be applied to posterior regularization for regulating other statistical models.

[Semantics] [ID:L28]

De-Conflated Semantic Representations
Mohammad Taher Pilehvar and Nigel Collier

One major deficiency of most semantic representation techniques is that they usually model a word type as a single point in the semantic space, hence conflating all the meanings that the word can have. Addressing this issue by learning distinct representations for individual meanings of words has been the subject of several research studies in the past few years. However, the generated sense representations are either not linked to any sense inventory or are unreliable for infrequent word senses. We propose a technique that tackles these problems by de-conflating the representations of words based on the deep knowledge that can be derived from a semantic network. Our approach provides multiple advantages in comparison to the previous approaches, including its high coverage and the ability to generate accurate representations even for infrequent word senses. We carry out evaluations on six datasets across two semantic similarity tasks and report state-of-the-art results on most of them.

[Semantics] [ID:L29]

Improving Sparse Word Representations with Distributional Inference for Semantic Composition
Thomas Kober, Julie Weeds, Jeremy Reffin, and David Weir

Distributional models are derived from co-occurrences in a corpus, where only a small proportion of all possible plausible co-occurrences will be observed. This results in a very sparse vector space, requiring a mechanism for inferring missing knowledge. Most methods face this challenge in ways that render the resulting word representations uninterpretable, with the consequence that semantic composition becomes hard to model. In this paper we explore an alternative which involves explicitly inferring unobserved co-occurrences using the distributional neighbourhood. We show that distributional inference improves sparse word representations on several word similarity benchmarks and demonstrate that our model is competitive with the state-of-the-art for adjective-noun, noun-noun and verb-object compositions while being fully interpretable.

[Semantics] [ID:L30]

Modelling Interaction of Sentence Pair with Coupled-LSTMs
Pengfei Liu, Xipeng Qiu, Yaqian Zhou, Jifan Chen, and Xuanjing Huang

Recently, there is rising interest in modelling the interactions of two sentences with deep neural networks. However, most of the existing methods encode two sequences with separate encoders, in which a sentence is encoded with little or no information from the other sentence. In this paper, we propose a deep architecture to model the strong interaction of sentence pair with two coupled-LSTMs. Specifically, we introduce two coupled ways to model the interdependences of two LSTMs, coupling the local contextualized interactions of two sentences. We then aggregate these interactions and use a dynamic pooling to select the most informative features. Experiments on two very large datasets demonstrate the efficacy of our proposed architectures.

[Semantics] [ID:L31]

Universal Decompositional Semantics on Universal Dependencies
Aaron Steven White, Drew Reisinger, Keisuke Sakaguchi, Tim Vieira, Sheng Zhang, Rachel Rudinger, Kyle Rawlins, and Benjamin Van Durme

We present a framework for augmenting data sets from the Universal Dependencies project with Universal Decompositional Semantics. Where the Universal Dependencies project aims to provide a syntactic dependency annotation standard that can be used consistently across many languages as well as a collection of corpora that use that standard, our extension has similar aims for semantic annotation. We describe results from annotating the English Universal Dependencies treebank, dealing with word senses, semantic roles, and event properties.

[Social Media & Computational Social Science] [ID:L32]

Friends with Motives: Using Text to Infer Influence on SCOTUS
Yanchuan Sim, Bryan Routledge, and Noah A. Smith

We present a probabilistic model of the influence of language on the behavior of the U.S.~Supreme Court, specifically influence of amicus briefs on Court decisions and opinions. The approach assumes that amici are rational, utility-maximizing agents who try to win votes or affect the language of court opinions. Our model leads to improved predictions of justices' votes and perplexity of opinion language. It is amenable to inspection, allowing us to explore inferences about the persuasiveness of different amici and influenceability of different justices; these are consistent with earlier findings.

[Syntax & Morphology] [ID:L33]
Verb Phrase Ellipsis Resolution Using Discriminative and Margin-Infused Algorithms
Kian Kenyon-Dean, Jackie Chi Kit Cheung, and Doina Precup

Verb Phrase Ellipsis (VPE) is an anaphoric construction in which a verb phrase has been elided. It occurs frequently in dialogue and informal conversational settings, but despite its evident impact on event coreference resolution and extraction, there has been relatively little work on computational methods for identifying and resolving VPE. Here, we present a novel approach to detecting and resolving VPE by using supervised discriminative machine learning techniques trained on features extracted from an automatically parsed, publicly available dataset. Our approach yields state-of-the-art results for VPE detection by improving F1 score by over 11%; additionally, we explore an approach to antecedent identification that uses the Margin-Infused-Relaxed-Algorithm, which shows promising results.

[Syntax & Morphology] [ID:L34]
Distilling an Ensemble of Greedy Dependency Parsers into One MST Parser
Adhiguna Kuncoro, Miguel Ballesteros, Lingpeng Kong, Chris Dyer, and Noah A. Smith

We introduce two first-order graph-based dependency parsers achieving a new state of the art. The first is a consensus parser built from an ensemble of independently trained greedy LSTM transition-based parsers with different random initializations. We cast this approach as minimum Bayes risk decoding (under the Hamming cost) and argue that weaker consensus within the ensemble is a useful signal of difficulty or ambiguity. The second parser is a "distillation" of the ensemble into a single model. We train the distillation parser using a structured hinge loss objective with a novel cost that incorporates ensemble uncertainty estimates for each possible attachment, thereby avoiding the intractable cross-entropy computations required by applying standard distillation objectives to problems with structured outputs. The first-order distillation parser matches or surpasses the state of the art on English, Chinese, and German.

[Syntax & Morphology] [ID:L35]
LSTM Shift-Reduce CCG Parsing
Wenduan Xu

We describe a neural shift-reduce parsing model for CCG, factored into four unidirectional LSTMs and one bidirectional LSTM. This factorization allows the linearization of the complete parsing history, and results in a highly accurate greedy parser that outperforms all previous beam-search shift-reduce parsers for CCG. By further deriving a globally optimized model using a task-based loss, we improve over the state of the art by up to 2.67% labeled F1.

[Syntax & Morphology] [ID:L36]
An Evaluation of Parser Robustness for Ungrammatical Sentences
Homa B. Hashemi and Rebecca Hwa

For many NLP applications that require a parser, the sentences of interest may not be well-formed. If the parser can overlook problems such as grammar mistakes and produce a parse tree that closely resembles the correct analysis for the intended sentence, we say that the parser is robust. This paper compares the performances of eight state-of-the-art dependency parsers on two domains of ungrammatical sentences: learner English and machine translation outputs. We have developed an evaluation metric and conducted a suite of experiments. Our analyses may help practitioners to choose an appropriate parser for their tasks, and help developers to improve parser robustness against ungrammatical sentences.

[Syntax & Morphology] [ID:L37]
Neural Shift-Reduce CCG Semantic Parsing
Dipendra Kumar Misra and Yoav Artzi

We present a shift-reduce CCG semantic parser. Our parser uses a neural network architecture that balances model capacity and computational cost. We train by transferring a model from a computationally expensive log-linear CKY parser. Our learner addresses two challenges: selecting the best parse for

learning when the CKY parser generates multiple correct trees, and learning from partial derivations when the CKY parser fails to parse. We evaluate on AMR parsing. Our parser performs comparably to the CKY parser, while doing significantly fewer operations. We also present results for greedy semantic parsing with a relatively small drop in performance.

[Syntax & Morphology] [ID:L38]

Syntactic Parsing of Web Queries

Xiangyan Sun, Haixun Wang, Yanghua Xiao, and Zhongyuan Wang

Syntactic parsing of web queries is important for query understanding. However, web queries usually do not observe the grammar of a written language, and no labeled syntactic trees for web queries are available. In this paper, we focus on a query's clicked sentence, i.e., a well-formed sentence that i) contains all the tokens of the query, and ii) appears in the query's top clicked web pages. We argue such sentences are semantically consistent with the query. We introduce algorithms to derive a query's syntactic structure from the dependency trees of its clicked sentences. This gives us a web query treebank without manual labeling. We then train a dependency parser on the treebank. Our model achieves much better UAS (0.86) and LAS (0.80) scores than state-of-the-art parsers on web queries.

[Summarization] [ID:L39]

Unsupervised Text Recap Extraction for TV Series

Hongliang Yu, Shikun Zhang, and Louis-Philippe Morency

Sequences found at the beginning of TV shows help the audience absorb the essence of previous episodes, and grab their attention with upcoming plots. In this paper, we propose a novel task, text recap extraction. Compared with conventional summarization, text recap extraction captures the duality of summarization and plot contingency between adjacent episodes. We present a new dataset, TVRecap, for text recap extraction on TV shows. We propose an unsupervised model that identifies text recaps based on plot descriptions. We introduce two contingency factors, concept coverage and sparse reconstruction, that encourage recaps to prompt the upcoming story development. We also propose a multi-view extension of our model which can incorporate dialogues and synopses. We conduct extensive experiments on TVRecap, and conclude that our model outperforms summarization approaches.

[Text Mining & Applications] [ID:L40]

On- and Off-Topic Classification and Semantic Annotation of User-Generated Software Requirements

Markus Dollmann and Michaela Geierhos

Users prefer natural language software requirements because of their usability and accessibility. When they describe their wishes for software development, they often provide off-topic information. We therefore present REaCT, an automated approach for identifying and semantically annotating the on-topic parts of requirement descriptions. It is designed to support requirement engineers in the elicitation process on detecting and analyzing requirements in user-generated content. Since no lexical resources with domain-specific information about requirements are available, we created a corpus of requirements written in controlled language by instructed users and uncontrolled language by uninstructed users. We annotated these requirements regarding predicate-argument structures, conditions, priorities, motivations and semantic roles and used this information to train classifiers for information extraction purposes. REaCT achieves an accuracy of 92% for the on- and off-topic classification task and an F1-measure of 72% for the semantic annotation.

[Text Mining & Applications] [ID:L41]

Deceptive Review Spam Detection via Exploiting Task Relatedness and Unlabeled Data

Zhen Hai, Peilin Zhao, Peng Cheng, Peng Yang, Xiao-Li Li, and Guangxia Li

Existing work on detecting deceptive reviews primarily focuses on feature engineering and applies off-the-shelf supervised classification algorithms to the problem. Then, one real challenge would be to manually recognize plentiful ground truth spam review data for model building, which is rather difficult and often requires domain expertise in practice. In this paper, we propose to exploit the relatedness of multiple review spam detection tasks and readily available unlabeled data to address the scarcity of labeled opinion spam data. We first develop a multi-task learning method based on logistic regression (MTL-LR), which can boost the learning for a task by sharing the knowledge contained in the training signals of other related tasks. To leverage the unlabeled data, we introduce a graph Laplacian regularizer into each base model. We then propose a novel semi-supervised multi-task learning method via Laplacian regularized logistic regression (SMTL-LLR) to further improve the review spam detection performance. We also develop a stochastic alternating method to cope with the optimization for SMTL-LLR. Experimental results on real-world review data demonstrate the benefit of SMTL-LLR over several well-established baseline methods.

[Text Mining & Applications] [ID:L42]

Regularizing Text Categorization with Clusters of Words

Konstantinos Skianis, Francois Rousseau, and Michalis Vazirgiannis

Regularization is a critical step in supervised learning to not only address overfitting, but also to take into account any prior knowledge we may have on the features and their dependence. In this paper, we explore state-of-the-art structured regularizers and we propose novel ones based on clusters of words from LSI topics, word2vec embeddings and graph-of-words document representation. We show that our proposed regularizers are faster than the state-of-the-art ones and still improve text classification accuracy.

[Text Mining & Applications] [ID:L43]

Deep Reinforcement Learning with a Combinatorial Action Space for Predicting Popular Reddit Threads

Ji He, Mari Ostendorf, Xiaodong He, Jianshu Chen, Jianfeng Gao, Lihong Li, and Li Deng

We introduce an online popularity prediction and tracking task as a benchmark task for reinforcement learning with a combinatorial, natural language action space. A specified number of discussion threads predicted to be popular are recommended, chosen from a fixed window of recent comments to track. Novel deep reinforcement learning architectures are studied for effective modeling of the value function associated with actions comprised of interdependent sub-actions. The proposed model, which represents dependence between sub-actions through a bi-directional LSTM, gives the best performance across different experimental configurations and domains, and it also generalizes well with varying numbers of recommendation requests.

[Text Mining & Applications] [ID:L44]

Non-Literal Text Reuse in Historical Texts: An Approach to Identify Reuse Transformations and its Application to Bible Reuse

Maria Moritz, Andreas Wiederhold, Barbara Pavlek, Yuri Bizzoni, and Marco Büchler

Text reuse refers to citing, copying or alluding text excerpts from a text resource to a new context. While detecting reuse in contemporary languages is well supported—given extensive research, techniques, and corpora— automatically detecting historical text reuse is much more difficult. Corpora of historical languages are less documented and often encompass various genres, linguistic varieties, and topics. In fact, historical text reuse detection is much less understood and empirical studies are necessary to enable and improve its automation. We present a linguistic analysis of text reuse in two ancient data sets. We contribute an automated approach to analyze how an original text was transformed into its reuse, taking linguistic resources into account to understand how they help characterizing the transformation. It is complemented by a manual analysis of a subset of the reuse. Our results show the limitations of approaches focusing on literal reuse detection. Yet, linguistic resources can effectively support understanding the non-literal text reuse transformation process. Our results support practitioners and researchers working on understanding and detecting historical reuse.

[Text Mining & Applications] [ID:L45]

A Graph Degeneracy-based Approach to Keyword Extraction

Antoine Tixier, Fragkiskos Malliaros, and Michalis Vazirgiannis

We operate a change of paradigm and hypothesize that keywords are more likely to be found among influential nodes of a graph-of-words rather than among its nodes high on eigenvector-related centrality measures. To test this hypothesis, we introduce unsupervised techniques that capitalize on graph degeneracy. Our methods strongly and significantly outperform all baselines on two datasets (short and medium size documents), and reach best performance on the third one (long documents).

[Text Mining & Applications] [ID:L46]

Predicting the Relative Difficulty of Single Sentences With and Without Surrounding Context

Elliot Schumacher, Maxine Eskenazi, Gwen Frishkoff, and Kevyn Collins-Thompson

The problem of accurately predicting relative reading difficulty across a set of sentences arises in a number of important natural language applications, such as finding and curating effective usage examples for intelligent language tutoring systems. Yet while significant research has explored document- and passage-level reading difficulty, the special challenges involved in assessing aspects of readability for single sentences have received much less attention, particularly when considering the role of surrounding passages. We introduce and evaluate a novel approach for estimating the relative reading difficulty of a set of sentences, with and without surrounding context. Using different sets of lexical and grammatical features, we explore models for predicting pairwise relative difficulty using logistic regression, and examine rankings generated by aggregating pairwise difficulty labels using a Bayesian rating system to

form a final ranking. We also compare rankings derived for sentences assessed with and without context, and find that contextual features can help predict differences in relative difficulty judgments across these two conditions.

[Text Mining & Applications] [ID:L47]

A Neural Approach to Automated Essay Scoring

Kaveh Taghipour and Hwee Tou Ng

Traditional automated essay scoring systems rely on carefully designed features to evaluate and score essays. The performance of such systems is tightly bound to the quality of the underlying features. However, it is laborious to manually design the most informative features for such a system. In this paper, we develop an approach based on recurrent neural networks to learn the relation between an essay and its assigned score, without any feature engineering. We explore several neural network models for the task of automated essay scoring and perform some analysis to get some insights of the models. The results show that our best system, which is based on long short-term memory networks, outperforms a strong baseline by 5.6% in terms of quadratic weighted Kappa, without requiring any feature engineering.

[Text Mining & Applications] [ID:L48]

Non-uniform Language Detection in Technical Writing

Weibo Wang, Abidalrahman Moh'd, Aminul Islam, Axel Soto, and Evangelos Milios

Technical writing in professional environments, such as user manual authoring, requires the use of uniform language. Non-uniform language detection is a novel task, which aims to guarantee the consistency for technical writing by detecting sentences in a document that are intended to have the same meaning within a similar context but use different words or writing style. This paper proposes an approach that utilizes text similarity algorithms at lexical, syntactic, semantic and pragmatic levels. Different features are extracted and integrated by applying a machine learning classification method. We tested our method using smart phone user manuals, and compared its performance against the state-of-the-art methods in a related area. The experiments demonstrate that our approach achieves the upper bound performance for this task.

[Text Mining & Applications] [ID:L49]

Adapting Grammatical Error Correction Based on the Native Language of Writers with Neural Network Joint Models

Shamil Chollampatt, Duc Tam Hoang, and Hwee Tou Ng

An important aspect for the task of grammatical error correction (GEC) that has not yet been adequately explored is adaptation based on the native language (L1) of writers, despite the marked influences of L1 on second language (L2) writing. In this paper, we adapt a neural network joint model (NNJM) using L1-specific learner text and integrate it into a statistical machine translation (SMT) based GEC system. Specifically, we train an NNJM on general learner text (not L1-specific) and subsequently train on L1-specific data using a Kullback-Leibler divergence regularized objective function in order to preserve generalization of the model. We incorporate this adapted NNJM as a feature in an SMT-based English GEC system and show that adaptation achieves significant F0.5 score gains on English texts written by L1 Chinese, Russian, and Spanish writers.

[Machine Translation] [ID:S01]

Orthographic Syllable as basic unit for SMT between Related Languages

Anoop Kunchukuttan and Pushpak Bhattacharyya

We explore the use of the orthographic syllable, a variable-length consonant-vowel sequence, as a basic unit of translation between related languages which use abugida or alphabetic scripts. We show that orthographic syllable level translation significantly outperforms models trained over other basic units (word, morpheme and character) when training over small parallel corpora.

[Text Mining & Applications] [ID:S02]

Neural Generation of Regular Expressions from Natural Language with Minimal Domain Knowledge

Nicholas Locascio, Karthik Narasimhan, Eduardo De Leon, Nate Kushman, and Regina Barzilay

This paper explores the task of translating natural language queries into regular expressions which embody their meaning. In contrast to prior work, the proposed neural model does not utilize domain-specific crafting, learning to translate directly from a parallel corpus. To fully explore the potential of neural models, we propose a methodology for collecting a large corpus of regular expression, natural

language pairs. Our resulting model achieves a performance gain of 19.6% over previous state-of-the-art models.

[Information Extraction] [ID:S03]

Supervised Keyphrase Extraction as Positive Unlabeled Learning

Lucas Sterckx, Cornelia Caragea, Thomas Demeester, and Chris Develder

The problem of noisy and unbalanced training data for supervised keyphrase extraction results from the subjectivity of keyphrase assignment, which we quantify by crowdsourcing keyphrases for news and fashion magazine articles with many annotators per document. We show that annotators exhibit substantial disagreement, meaning that single annotator data could lead to very different training sets for supervised keyphrase extractors. Thus, annotations from single authors or readers lead to noisy training data and poor extraction performance of the resulting supervised extractor. We provide a simple but effective solution to still work with such data by reweighting the importance of unlabeled candidate phrases in a two stage Positive Unlabeled Learning setting. We show that performance of trained keyphrase extractors approximates a classifier trained on articles labeled by multiple annotators, leading to higher average F1 scores and better rankings of keyphrases. We apply this strategy to a variety of test collections from different backgrounds and show improvements over strong baseline models.

[Information Extraction] [ID:S04]

Learning to Answer Questions from Wikipedia Infoboxes

Alvaro Morales, Varot Premtoon, Cordelia Avery, Sue Felshin, and Boris Katz

A natural language interface to answers on the Web can help us access information more efficiently. We start with an interesting source of information—infoboxes in Wikipedia that summarize factoid knowledge—and develop a comprehensive approach to answering questions with high precision. We first build a system to access data in infoboxes in a structured manner. We use our system to construct a crowdsourced dataset of over 15,000 high-quality, diverse questions. With these questions, we train a convolutional neural network model that outperforms models that achieve top results in similar answer selection tasks.

[Information Extraction] [ID:S05]

Timeline extraction using distant supervision and joint inference

Savelie Cornegruta and Andreas Vlachos

In timeline extraction the goal is to order all the events in which a target entity is involved in a timeline. Due to the lack of explicitly annotated data, previous work is primarily rule-based and uses pre-trained temporal linking systems. In this work, we propose a distantly supervised approach by heuristically aligning timelines with documents. The noisy training data created allows us to learn models that anchor events to temporal expressions and entities; during testing, the predictions of these models are combined to produce the timeline. Furthermore, we show how to improve performance using joint inference. In experiments in the SemEval-2015 TimeLine task we show that our distantly supervised approach matches the state-of-the-art performance while joint inference further improves on it by 3.2 F-score points.

[Information Extraction] [ID:S06]

Combining Supervised and Unsupervised Enembles for Knowledge Base Population

Nazneen Fatema Rajani and Raymond Mooney

We combine supervised and unsupervised methods to ensemble systems for two popular Knowledge Base Population (KBP) tasks, Cold Start Slot Filling (CSSF) and Tri-lingual Entity Discovery and Linking (TEDL). We demonstrate that it outperforms the best system for both tasks in the 2015 competition, several ensembling baselines, as well as a state-of-the-art stacking approach. The success of our technique on two different and challenging problems demonstrates the power and generality of our combined approach to ensembling.

[Language and Vision] [ID:S07]

Character Sequence Models for Colorful Words

Kazuya Kawakami, Chris Dyer, Bryan Routledge, and Noah A. Smith

We present a neural network architecture to predict a point in color space from the sequence of characters in the color's name. Using large scale color–name pairs obtained from an online color design forum, we evaluate our model on a "color Turing test" and find that, given a name, the colors predicted by our model are preferred by annotators to color names created by humans. Our datasets and demo system are available online at http://colorlab.us.

[Language and Vision] [ID:S08]

Analyzing the Behavior of Visual Question Answering Models

Aishwarya Agrawal, Dhruv Batra, and Devi Parikh

Recently, a number of deep-learning based models have been proposed for the task of Visual Question Answering (VQA). The performance of most models is clustered around 60-70%. In this paper we propose systematic methods to analyze the behavior of these models as a first step towards recognizing their strengths and weaknesses, and identifying the most fruitful directions for progress. We analyze two models, one each from two major classes of VQA models – with-attention and without-attention and show the similarities and differences in the behavior of these models. We also analyze the winning entry of the VQA Challenge 2016. Our behavior analysis reveals that despite recent progress, today's VQA models are "myopic" (tend to fail on sufficiently novel instances), often "jump to conclusions" (converge on a predicted answer after 'listening' to just half the question), and are "stubborn" (do not change their answers across images).

[Language and Vision] [ID:S09]

Improving LSTM-based Video Description with Linguistic Knowledge Mined from Text

Subhashini Venugopalan, Lisa Anne Hendricks, Raymond Mooney, and Kate Saenko

This paper investigates how linguistic knowledge mined from large text corpora can aid the generation of natural language descriptions of videos. Specifically, we integrate both a neural language model and distributional semantics trained on large text corpora into a recent LSTM-based architecture for video description. We evaluate our approach on a collection of Youtube videos as well as two large movie description datasets showing significant improvements in grammaticality while modestly improving descriptive quality.

[Semantics] [ID:S10]

Representing Verbs with Rich Contexts: an Evaluation on Verb Similarity

Emmanuele Chersoni, Enrico Santus, Alessandro Lenci, Philippe Blache, and Chu-Ren Huang

Several studies on sentence processing suggest that the mental lexicon keeps track of the mutual expectations between words. Current DSMs, however, represent context words as separate features, thereby loosing important information for word expectations, such as word interrelations. In this paper, we present a DSM that addresses this issue by defining verb contexts as joint syntactic dependencies. We test our representation in a verb similarity task on two datasets, showing that joint contexts achieve performances comparable to single dependencies or even better. Moreover, they are able to overcome the data sparsity problem of joint feature spaces, in spite of the limited size of our training corpus.

[Machine Learning] [ID:S11]

Speed-Accuracy Tradeoffs in Tagging with Variable-Order CRFs and Structured Sparsity

Tim Vieira, Ryan Cotterell, and Jason Eisner

We propose a method for learning the structure of variable-order CRFs, a more flexible variant of higher-order linear-chain CRFs. Variable-order CRFs achieve faster inference by including features for only some of the tag n-grams. Our learning method discovers the useful higher-order features at the same time as it trains their weights, by maximizing an objective that combines log-likelihood with a structured-sparsity regularizer. An active-set outer loop allows the feature set to grow as far as needed. On part-of-speech tagging in 5 randomly chosen languages from the Universal Dependencies dataset, our method of shrinking the model achieved a 2-6x speedup over a baseline, with no significant drop in accuracy.

[Machine Learning] [ID:S12]

Learning Robust Representations of Text

Yitong Li, Trevor Cohn, and Timothy Baldwin

Deep neural networks have achieved remarkable results across many language processing tasks, however these methods are highly sensitive to noise and adversarial attacks. We present a regularization based method for limiting network sensitivity to its inputs, inspired by ideas from computer vision, thus learning models that are more robust. Empirical evaluation over a range of sentiment datasets with a convolutional neural network shows that, compared to a baseline model and the dropout method, our method achieves superior performance over noisy inputs and out-of-domain data.

[Machine Learning] [ID:S13]

Modified Dirichlet Distribution: Allowing Negative Parameters to Induce Stronger Sparsity
Kewei Tu

The Dirichlet distribution (Dir) is one of the most widely used prior distributions in statistical approaches to natural language processing. The parameters of Dir are required to be positive, which significantly limits its strength as a sparsity prior. In this paper, we propose a simple modification to the Dirichlet distribution that allows the parameters to be negative. Our modified Dirichlet distribution (mDir) not only induces much stronger sparsity, but also simultaneously performs smoothing. mDir is still conjugate to the multinomial distribution, which simplifies posterior inference. We introduce two simple and efficient algorithms for finding the mode of mDir. Our experiments on learning Gaussian mixtures and unsupervised dependency parsing demonstrate the advantage of mDir over Dir.

[Machine Learning] [ID:S14]

Gated Word-Character Recurrent Language Model
Yasumasa Miyamoto and Kyunghyun Cho

We introduce a recurrent neural network language model (RNN-LM) with long short-term memory (LSTM) units that utilizes both character-level and word-level inputs. Our model has a gate that adaptively finds the optimal mixture of the character-level and word-level inputs. The gate creates the final vector representation of a word by combining two distinct representations of the word. The character-level inputs are converted into vector representations of words using a bidirectional LSTM. The word-level inputs are projected into another high-dimensional space by a word lookup table. The final vector representations of words are used in the LSTM language model which predicts the next word given all the preceding words. Our model with the gating mechanism effectively utilizes the character-level inputs for rare and out-of vocabulary words and outperforms word-level language models on several English corpora.

[Syntax & Morphology] [ID:S15]

Unsupervised Word Alignment by Agreement Under ITG Constraint
Hidetaka Kamigaito, Akihiro Tamura, Hiroya Takamura, Manabu Okumura, and Eiichiro Sumita

We propose a novel unsupervised word alignment method that uses a constraint based on Inversion Transduction Grammar (ITG) parse trees to jointly unify two directional models. Previous agreement methods are not helpful for locating alignments with long distances because they do not use any syntactic structures. In contrast, the proposed method symmetrizes alignments in consideration of their structural coherence by using the ITG constraint softly in the posterior regularization framework (Ganchev et al., 2010). The ITG constraint is also compatible with word alignments that are not covered by ITG parse trees. Hence, the proposed method is robust to ITG parse errors compared to other alignment methods that directly use an ITG model. Compared to the baseline agreement method (Ganchev et al., 2010), the experimental results show that the proposed method significantly improves alignment performance regarding Japanese-English KFTT and BTEC corpus, and machine translation performance on the Japanese-English IWSLT 2007 corpus.

[Syntax & Morphology] [ID:S16]

Training with Exploration Improves a Greedy Stack LSTM Parser
Miguel Ballesteros, Yoav Goldberg, Chris Dyer, and Noah A. Smith

We adapt the greedy stack LSTM dependency parser of Dyer et al. (2015) to support a training-with-exploration procedure using dynamic oracles (Goldberg and Nivre, 2013) instead of assuming an error-free action history. This form of training, which accounts for model predictions at training time, improves parsing accuracies. We discuss some modifications needed in order to get training with exploration to work well for a probabilistic neural-network dependency parser.

[Semantics] [ID:S17]

Capturing Argument Relationship for Chinese Semantic Role Labeling
Lei Sha, Sujian Li, Baobao Chang, Zhifang Sui, and Tingsong Jiang

In this paper, we capture the argument relationships for Chinese semantic role labeling task, and improve the task's performance with the help of argument relationships. We split the relationship between two candidate arguments into two categories: (1) Compatible arguments: if one candidate argument belongs to a given predicate, then the other is more likely to belong to the same predicate; (2) Incompatible arguments: if one candidate argument belongs to a given predicate, then the other is less likely to belong to the same predicate. However, previous works did not explicitly model argument relationships. We use a simple maximum entropy classifier to capture the two categories of argument relationships and test its performance on the Chinese Proposition Bank (CPB). The experiments show that argument relationships is effective in Chinese semantic role labeling task.

[Semantics] [ID:S18]

BrainBench: A Brain-Image Test Suite for Distributional Semantic Models

Haoyan Xu, Brian Murphy, and Alona Fyshe

The brain is the locus of our language ability, and so brain images can be used to ground linguistic theories. Here we introduce BrainBench, a lightweight system for testing distributional models of word semantics. We compare the performance of several models, and show that the performance on brain-image tasks differs from the performance on behavioral tasks. We release our benchmark test as part of a web service.

[Semantics] [ID:S19]

Evaluating Induced CCG Parsers on Grounded Semantic Parsing

Yonatan Bisk, Siva Reddy, John Blitzer, Julia Hockenmaier, and Mark Steedman

We compare the effectiveness of four different syntactic CCG parsers for a semantic slot-filling task to explore how much syntactic supervision is required for downstream semantic analysis. This extrinsic, task-based evaluation also provides a unique window into the semantics captured (or missed) by unsupervised grammar induction systems.

[Semantics] [ID:S20]

Vector-space models for PPDB paraphrase ranking in context

Marianna Apidianaki

The PPDB is an automatically built database which contains millions of paraphrases in different languages. Paraphrases in this resource are associated with features that serve to their ranking and reflect paraphrase quality. This context-unaware ranking captures the semantic similarity of paraphrases but cannot serve to estimate their adequacy in specific contexts. We propose to use vector-space semantic models for selecting PPDB paraphrases that preserve the meaning of specific text fragments. This is the first work that addresses the substitutability of PPDB paraphrases in context. We show that vector-space models of meaning can be successfully applied to this task and increase the benefit brought by the use of the PPDB resource in applications.

[Sentiment Analysis] [ID:S21]

Interpreting Neural Networks to Improve Politeness Comprehension

Malika Aubakirova and Mohit Bansal

We present an interpretable neural network approach to predicting and understanding politeness in natural language requests. Our models are based on simple convolutional neural networks directly on raw text, avoiding any manual identification of complex sentiment or syntactic features, while performing better than such feature-based models from previous work. More importantly, we use the challenging task of politeness prediction as a testbed to next present a much-needed understanding of what these successful networks are actually learning. For this, we present several network visualizations based on activation clusters, first derivative saliency, and embedding space transformations, helping us automatically identify several subtle linguistics markers of politeness theories. Further, this analysis reveals multiple novel, high-scoring politeness strategies which, when added back as new features, reduce the accuracy gap between the original featurized system and the neural model, thus providing a clear quantitative interpretation of the success of these neural networks.

[Sentiment Analysis] [ID:S22]

Does 'well-being' translate on Twitter?

Laura Smith, Salvatore Giorgi, Rishi Solanki, Johannes Eichstaedt, H. Andrew Schwartz, Muhammad Abdul-Mageed, Anneke Buffone, and Lyle Ungar

We investigate whether psychological well-being translates across English and Spanish Twitter, by building and comparing source language and automatically translated weighted lexica in English and Spanish. We find that the source language models perform substantially better than the machine translated versions. Moreover, manually correcting translation errors does not improve model performance, suggesting that meaningful cultural information is being lost in translation. Further work is needed to clarify when automatic translation of well-being lexica is effective and how it can be improved for cross-cultural analysis.

[Social Media & Computational Social Science] [ID:S23]

Beyond Canonical Texts: A Computational Analysis of Fanfiction

Smitha Milli and David Bamman

While much computational work on fiction has focused on works in the literary canon, user-created fanfiction presents a unique opportunity to study an ecosystem of literary production and consumption,

embodying qualities both of large-scale literary data (55 billion tokens) and also a social network (with over 2 million users). We present several empirical analyses of this data in order to illustrate the range of affordances it presents to research in NLP, computational social science and the digital humanities. We find that fanfiction deprioritizes main protagonists in comparison to canonical texts, has a statistically significant difference in attention allocated to female characters, and offers promise for predicting reader reactions to stories.

[Social Media & Computational Social Science] [ID:S24]

Using Syntactic and Semantic Context to Explore Psychodemographic Differences in Self-reference

Masoud Rouhizadeh, Lyle Ungar, Anneke Buffone, and H. Andrew Schwartz

Psychological analysis of language has repeatedly shown that an individual's rate of mentioning 1st person singular pronouns predicts a wealth of important demographic and psychological factors. However, these analyses are performed out of context — syntactic and semantic — which may change the magnitude or even direction of such relationships. In this paper, we put "pronouns in their context", exploring the relationship between self-reference and age, gender, and depression depending on syntactic position and verbal governor. We find that pronouns are overall more predictive when taking dependency relations and verb semantic categories into account, and, the direction of the relationship can change depending on the semantic class of the verbal governor.

[Social Media & Computational Social Science] [ID:S25]

Learning to Identify Metaphors from a Corpus of Proverbs

Gözde Özbal, Carlo Strapparava, Serra Sinem Tekiroglu, and Daniele Pighin

In this paper, we experiment with a resource consisting of metaphorically annotated proverbs on the task of word-level metaphor recognition. We observe that existing feature sets do not perform well on this data. We design a novel set of features to better capture the peculiar nature of proverbs and we demonstrate that these new features are significantly more effective on the metaphorically dense proverb data.

[Social Media & Computational Social Science] [ID:S26]

An Embedding Model for Predicting Roll-Call Votes

Peter Kraft, Hirsh Jain, and Alexander M. Rush

We develop a novel embedding-based model for predicting legislative roll-call votes from bill text. The model introduces multidimensional *ideal vectors* for legislators as an alternative to single dimensional ideal point models for quantitatively analyzing roll-call data. These vectors are learned to correspond with pre-trained word embeddings which allows us to analyze which features in a bill text are most predictive of political support. Our model is quite simple, while at the same time allowing us to successfully predict legislator votes on specific bills with higher accuracy than past methods.

[Spoken Language Processing] [ID:S27]

Natural Language Model Re-usability for Scaling to Different Domains

Young-Bum Kim, Alexandre Rochette, and Ruhi Sarikaya

Natural language understanding is the core of the human computer interactions. However, building new domains and tasks that need a separate set of models is a bottleneck for scaling to a large number of domains and experiences. In this paper, we propose a practical technique that addresses this issue in a web-scale language understanding system: Microsoft's personal digital assistant Cortana. The proposed technique uses a constrained decoding method with a universal slot tagging model sharing the same schema as the collection of slot taggers built for each domain. The proposed approach allows reusing of slots across different domains and tasks while achieving virtually the same performance as those slot taggers trained per domain fashion.

[Spoken Language Processing] [ID:S28]

Leveraging Sentence-level Information with Encoder LSTM for Semantic Slot Filling

Gakuto Kurata, Bing Xiang, Bowen Zhou, and Mo Yu

Recurrent Neural Network (RNN) and one of its specific architectures, Long Short-Term Memory (LSTM), have been widely used for sequence labeling. Explicitly modeling output label dependencies on top of RNN/LSTM is a widely-studied and effective extension. We propose another extension to incorporate the global information spanning over the whole input sequence. The proposed method, encoder-labeler LSTM, first encodes the whole input sequence into a fixed length vector with the encoder LSTM, and then uses this encoded vector as the initial state of another LSTM for sequence labeling. With this method, we

can predict the label sequence while taking the whole input sequence information into consideration. In the experiments of a slot filling task, which is an essential component of natural language understanding, with using the standard ATIS corpus, we achieved the state-of-the-art F1-score of 95.66%.

[Summarization] [ID:S29]

AMR-to-text generation as a Traveling Salesman Problem

Linfeng Song, Yue Zhang, Xiaochang Peng, Zhiguo Wang, and Daniel Gildea

The task of AMR-to-text generation is to gen- erate grammatical text that sustains the seman- tic meaning for a given AMR graph. We at- tack the task by first partitioning the AMR graph into smaller fragments, and then gener- ating the translation for each fragment, before finally deciding the order by solving an asym- metric generalized traveling salesman prob- lem (AGTSP). A Maximum Entropy classifier is trained to estimate the traveling costs, and a TSP solver is used to find the optimized solu- tion. The final model reports a BLEU score of 22.44 on the SemEval-2016 Task8 dataset.

[Text Mining & Applications] [ID:S30]

Learning to Capitalize with Character-Level Recurrent Neural Networks: An Empirical Study

Raymond Hendy Susanto, Hai Leong Chieu, and Wei Lu

In this paper, we investigate case restoration for text without case information. Previous such work operates at the word level. We propose an approach using character-level recurrent neural networks (RNN), which performs competitively compared to language modeling and conditional random fields (CRF) approaches. We further provide quantitative and qualitative analysis on how RNN helps improve truecasing.

[Text Mining & Applications] [ID:S31]

The Effects of the Content of FOMC Communications on US Treasury Rates

Christopher Rohlfs, Sunandan Chakraborty, and Lakshminarayanan Subramanian

This study measures the effects of Federal Open Market Committee text content on the direction of short- and medium-term interest rate movements. Because the words relevant to short- and medium-term interest rates differ, we apply a supervised approach to learn distinct sets of topics for each dependent variable being examined. We generate predictions with and without controlling for factors relevant to interest rate movements, and our prediction results average across multiple training-test splits. Using data from 1999-2016, we achieve 93% and 64% accuracy in predicting Target and Effective Federal Funds Rate movements and 38%-40% accuracy in predicting longer term Treasury Rate movements. We obtain lower but comparable accuracies after controlling for other macroeconomic and market factors.

[Text Mining & Applications] [ID:S32]

Learning to refine text based recommendations

Youyang Gu, Tao Lei, Regina Barzilay, and Tommi Jaakkola

We propose a text-based recommendation engine that utilizes recurrent neural networks to flexibly map textual input into continuous vector representations tailored to the recommendation task. Here, the text objects are documents such as Wikipedia articles or question and answer pairs. As neural models require substantial training time, we introduce a sequential component so as to quickly adjust the learned metric over objects as additional evidence accrues. We evaluate the approach on recommending Wikipedia descriptions of ingredients to their associated product categories. We also exemplify the sequential metric adjustment on retrieving similar Stack Exchange AskUbuntu questions.

[Text Mining & Applications] [ID:S33]

There's No Comparison: Reference-less Evaluation Metrics in Grammatical Error Correction

Courtney Napoles, Keisuke Sakaguchi, and Joel Tetreault

Current methods for automatically evaluating grammatical error correction (GEC) systems rely on gold-standard references. However, these methods suffer from penalizing grammatical edits that are correct but not in the gold standard. We show that reference-less grammaticality metrics correlate very strongly with human judgments and are competitive with the leading reference-based evaluation metrics. By interpolating both methods, we achieve state-of-the-art correlation with human judgments. Finally, we show that GEC metrics are much more reliable when they are calculated at the sentence level instead of the corpus level. We have set up a CodaLab site for benchmarking GEC output using a common dataset and different evaluation metrics.

6

Main Conference: Friday, November 4

Overview

07:30 – 17:30 **Registration Day 3** *Salon GHJ Prefunction*

08:00 – 09:00 **Morning Coffee** *Salon GHJ Prefunction*

09:00 – 10:00 **Session P3: Plenary Session III** *Salon FG*

09:00 – 10:00 *Invited Talk: You Talking to Me? Speech-based and multimodal approaches for human versus computer addressee detection (Andreas Stolcke)*

10:00 – 10:30 **Coffee Break** *Salon GHJ Prefunction*

10:30 – 12:10 **Session 7**

Dialogue Systems	Semantic Similarity	Dependency Parsing
Salon FG	*Salon J*	*Room 616 AB*

12:10 – 13:40 **Lunch**

13:40 – 15:25 **Session 8**

Short Paper Oral Session I	Short Paper Oral Session II	Short Paper Oral Session III
Salon FG	*Salon J*	*Room 616 AB*

15:25 – 15:50 **Coffee Break** *Salon GHJ Prefunction*

15:50 – 17:35 **Session P4: Best Paper Plenary Session** *Salon FG*

15:50 – 15:55 *Introduction to Best Papers (Program Chairs)*

15:55 – 16:20 Improving Information Extraction by Acquiring External Evidence with Reinforcement Learning (p. 112)
Karthik Narasimhan, Adam Yala, and Regina Barzilay

16:20 – 16:45 Global Neural CCG Parsing with Optimality Guarantees (p. 112)
Kenton Lee, Mike Lewis, and Luke Zettlemoyer

16:45 – 17:10 Learning a Lexicon and Translation Model from Phoneme Lattices (p. 112)
Oliver Adams, Graham Neubig, Trevor Cohn, Steven Bird, Quoc Truong Do, and Satoshi Nakamura

17:10 – 17:35 SQuAD: 100,000+ Questions for Machine Comprehension of Text (p. 112)
Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang

17:35 – 17:55 **Session P5: Closing Remarks Plenary Session** *Salon FG*

Invited Speaker: Andreas Stolcke

You Talking to Me? Speech-based and multimodal approaches for human versus computer addressee detection

Friday, November 4, 2016,

Abstract: As dialog systems become ubiquitous, we must learn how to detect when a system is spoken to, and avoid mistaking human-human speech as computer-directed input. In this talk I will discuss approaches to addressee detection in this human-human-machine dialog scenario, based on what is being said (lexical information), how it is being said (acoustic-prosodic properties), and non-speech multimodal and contextual information. I will present experimental results showing that a combination of these cues can be used effectively for human/computer address classification in several dialog scenarios.

Biography: Andreas Stolcke received a Ph.D. in computer science from the University of California at Berkeley. He was subsequently a Senior Research Engineer with the Speech Technology and Research Laboratory at SRI International, Menlo Park, CA, and is currently a Principal Researcher with the Speech and Dialog Research Group in the Microsoft Advanced Technology-Information Services group, working out of Mountain View, CA. His research interests include language modeling, speech recognition, speaker recognition, and speech understanding. He has published over 200 papers in these areas, as well as SRILM, a widely used open-source toolkit for statistical language modeling. He is a Fellow of the IEEE and of ISCA, the International Speech Communications Association.

Session 7 Overview

Track A	**Track B**	**Track C**	
Dialogue Systems Salon FG	*Semantic Similarity* Salon J	*Dependency Parsing* Room 616 AB	
How NOT To Evaluate Your Dialogue System: An Empirical Study of Unsupervised Evaluation Metrics for Dialogue Response Generation *Liu, Lowe, Serban, Noseworthy, Charlin, and Pineau*	Relations such as Hypernymy: Identifying and Exploiting Hearst Patterns in Distributional Vectors for Lexical Entailment *Roller and Erk*	Bi-directional Attention with Agreement for Dependency Parsing *Cheng, Fang, He, Gao, and Deng*	10:30
Addressee and Response Selection for Multi-Party Conversation *Ouchi and Tsuboi*	SimVerb-3500: A Large-Scale Evaluation Set of Verb Similarity *Gerz, Vulić, Hill, Reichart, and Korhonen*	The Galactic Dependencies Treebanks: Getting More Data by Synthesizing New Languages *Wang and Eisner*	10:55
Nonparametric Bayesian Models for Spoken Language Understanding *Wakabayashi, Takeuchi, Funakoshi, and Nakano*	POLY: Mining Relational Paraphrases from Multilingual Sentences *Grycner and Weikum*	Easy-First Dependency Parsing with Hierarchical Tree LSTMs *Kiperwasser and Goldberg*	11:20
Conditional Generation and Snapshot Learning in Neural Dialogue Systems *Wen, Gasic, Mrkšić, Rojas Barahona, Su, Ultes, Vandyke, and Young*	Exploiting Sentence Similarities for Better Alignments *Li and Srikumar*	Anchoring and Agreement in Syntactic Annotations *Berzak, Huang, Barbu, Korhonen, and Katz*	11:45

Session 7A: Dialogue Systems

Salon FG

How NOT To Evaluate Your Dialogue System: An Empirical Study of Unsupervised Evaluation Metrics for Dialogue Response Generation

Chia-Wei Liu, Ryan Lowe, Iulian Serban, Mike Noseworthy, Laurent Charlin, and Joelle Pineau 10:30–10:55

We investigate evaluation metrics for dialogue response generation systems where supervised labels, such as task completion, are not available. Recent works in response generation have adopted metrics from machine translation to compare a model's generated response to a single target response. We show that these metrics correlate very weakly with human judgements in the non-technical Twitter domain, and not at all in the technical Ubuntu domain. We provide quantitative and qualitative results highlighting specific weaknesses in existing metrics, and provide recommendations for future development of better automatic evaluation metrics for dialogue systems.

Addressee and Response Selection for Multi-Party Conversation

Hiroki Ouchi and Yuta Tsuboi 10:55–11:20

To create conversational systems working in actual situations, it is crucial to assume that they interact with multiple agents. In this work, we tackle addressee and response selection for multi-party conversation, in which systems are expected to select whom they address as well as what they say. The key challenge of this task is to jointly model who is talking about what in a previous context. For the joint modeling, we propose two modeling frameworks: 1) static modeling and 2) dynamic modeling. To show benchmark results of our frameworks, we created a multi-party conversation corpus. Our experiments on the dataset show that the recurrent neural network based models of our frameworks robustly predict addressees and responses in conversations with a large number of agents.

Nonparametric Bayesian Models for Spoken Language Understanding

Kei Wakabayashi, Johane Takeuchi, Kotaro Funakoshi, and Mikio Nakano 11:20–11:45

In this paper, we propose a new generative approach for semantic slot filling task in spoken language understanding using a nonparametric Bayesian formalism. Slot filling is typically formulated as a sequential labeling problem, which does not directly deal with the posterior distribution of possible slot values. We present a nonparametric Bayesian model involving the generation of arbitrary natural language phrases, which allows an explicit calculation of the distribution over an infinite set of slot values. We demonstrate that this approach significantly improves slot estimation accuracy compared to the existing sequential labeling algorithm.

Conditional Generation and Snapshot Learning in Neural Dialogue Systems

Tsung-Hsien Wen, Milica Gasic, Nikola Mrkšić, Lina M. Rojas Barahona, Pei-Hao Su, Stefan Ultes, David Vandyke, and Steve Young 11:45–12:10

Recently a variety of LSTM-based conditional language models (LM) have been applied across a range of language generation tasks. In this work we study various model architectures and different ways to represent and aggregate the source information in an end-to-end neural dialogue system framework. A method called snapshot learning is also proposed to facilitate learning from supervised sequential signals by applying a *companion* cross-entropy objective function to the conditioning vector. The experimental and analytical results demonstrate firstly that competition occurs between the conditioning vector and the LM, and the differing architectures provide different trade-offs between the two. Secondly, the discriminative power and transparency of the conditioning vector is key to providing both model interpretability and better performance. Thirdly, snapshot learning leads to consistent performance improvements independent of which architecture is used.

Session 7B: Semantic Similarity

Salon J

Relations such as Hypernymy: Identifying and Exploiting Hearst Patterns in Distributional Vectors for Lexical Entailment

Stephen Roller and Katrin Erk 10:30–10:55

We consider the task of predicting lexical entailment using distributional vectors. We perform a novel qualitative analysis of one existing model which was previously shown to only measure the prototypicality of word pairs. We find that the model strongly learns to identify hypernyms using Hearst patterns, which are well known to be predictive of lexical relations. We present a novel model which exploits this behavior as a method of feature extraction in an iterative procedure similar to Principal Component Analysis. Our model combines the extracted features with the strengths of other proposed models in the literature, and matches or outperforms prior work on multiple data sets.

SimVerb-3500: A Large-Scale Evaluation Set of Verb Similarity

Daniela Gerz, Ivan Vulić, Felix Hill, Roi Reichart, and Anna Korhonen 10:55–11:20

Verbs play a critical role in the meaning of sentences, but these ubiquitous words have received little attention in recent distributional semantics research. We introduce SimVerb-3500, an evaluation resource that provides human ratings for the similarity of 3,500 verb pairs. SimVerb-3500 covers all normed verb types from the USF free-association database, providing at least three examples for every VerbNet class. This broad coverage facilitates detailed analyses of how syntactic and semantic phenomena together influence human understanding of verb meaning. Further, with significantly larger development and test sets than existing benchmarks, SimVerb-3500 enables more robust evaluation of representation learning architectures and promotes the development of methods tailored to verbs. We hope that SimVerb-3500 will enable a richer understanding of the diversity and complexity of verb semantics and guide the development of systems that can effectively represent and interpret this meaning.

POLY: Mining Relational Paraphrases from Multilingual Sentences

Adam Grycner and Gerhard Weikum 11:20–11:45

Language resources that systematically organize paraphrases for binary relations are of great value for various NLP tasks and have recently been advanced in projects like PATTY, WiseNet and DEFIE. This paper presents a new method for building such a resource and the resource itself, called POLY. Starting with a very large collection of multilingual sentences parsed into triples of phrases, our method clusters relational phrases using probabilistic measures. We judiciously leverage fine-grained semantic typing of relational arguments for identifying synonymous phrases. The evaluation of POLY shows significant improvements in precision and recall over the prior works on PATTY and DEFIE. An extrinsic use case demonstrates the benefits of POLY for question answering.

Exploiting Sentence Similarities for Better Alignments

Tao Li and Vivek Srikumar 11:45–12:10

We study the problem of jointly aligning sentence constituents and predicting their similarities. While extensive sentence similarity data exists, manually generating reference alignments and labeling the similarities of the aligned chunks is comparatively onerous. This prompts the natural question of whether we can exploit easy-to-create sentence level data to train better aligners. In this paper, we present a model that learns to jointly align constituents of two sentences and also predict their similarities. By taking advantage of both sentence and constituent level data, we show that our model achieves state-of-the-art performance at predicting alignments and constituent similarities.

Session 7C: Dependency Parsing

Room 616 AB

Bi-directional Attention with Agreement for Dependency Parsing

Hao Cheng, Hao Fang, Xiaodong He, Jianfeng Gao, and Li Deng 10:30–10:55

We develop a novel bi-directional attention model for dependency parsing, which learns to agree on headword predictions from the forward and backward parsing directions. The parsing procedure for each direction is formulated as sequentially querying the memory component that stores continuous headword embeddings. The proposed parser makes use of *soft* headword embeddings, allowing the model to implicitly capture high-order parsing history without dramatically increasing the computational complexity. We conduct experiments on English, Chinese, and 12 other languages from the CoNLL 2006 shared task, showing that the proposed model achieves state-of-the-art unlabeled attachment scores on 6 languages.

The Galactic Dependencies Treebanks: Getting More Data by Synthesizing New Languages

Dingquan Wang and Jason Eisner 10:55–11:20

We release Galactic Dependencies 1.0—a large set of synthetic languages not found on Earth, but annotated in Universal Dependencies format. This new resource aims to pro-vide training and development data for NLP methods that aim to adapt to unfamiliar languages. Each synthetic treebank is produced from a real treebank by stochastically permuting the dependents of nouns and/or verbs to match the word order of other real languages. We discuss the usefulness, realism, parsability, perplexity, and diversity of the synthetic languages. As a simple demonstration of the use of Galactic Dependencies, we consider single-source transfer, which attempts to parse a real target language using a parser trained on a "nearby" source language. We find that including synthetic source languages somewhat increases the diversity of the source pool, which significantly improves results for most target languages.

Easy-First Dependency Parsing with Hierarchical Tree LSTMs

Eliyahu Kiperwasser and Yoav Goldberg 11:20–11:45

We suggest a compositional vector representation of parse trees that relies on a recursive combination of recurrent-neural network encoders. To demonstrate its effectiveness, we use the representation as the backbone of a greedy, bottom-up dependency parser, achieving very strong accuracies for English and Chinese, without relying on external word embeddings. The parser's implementation is available for download at the first author's webpage.

Anchoring and Agreement in Syntactic Annotations

Yevgeni Berzak, Yan Huang, Andrei Barbu, Anna Korhonen, and Boris Katz 11:45–12:10

We present a study on two key characteristics of human syntactic annotations: anchoring and agreement. Anchoring is a well known cognitive bias in human decision making, where judgments are drawn towards pre-existing values. We study the influence of anchoring on a standard approach to creation of syntactic resources where syntactic annotations are obtained via human editing of tagger and parser output. Our experiments demonstrate a clear anchoring effect and reveal unwanted consequences, including overestimation of parsing performance and lower quality of annotations in comparison with human-based annotations. Using sentences from the Penn Treebank WSJ, we also report systematically obtained inter-annotator agreement estimates for English dependency parsing. Our agreement results control for parser bias, and are consequential in that they are on par with state of the art parsing performance for English newswire. We discuss the impact of our findings on strategies for future annotation efforts and parser evaluations.

Session 8 Overview

Track A	**Track B**	**Track C**	
Short Paper Oral Session I Salon FG	*Short Paper Oral Session II* Salon J	*Short Paper Oral Session III* Room 616 AB	
Tense Manages to Predict Implicative Behavior in Verbs *Pavlick and Callison-Burch*	Insertion Position Selection Model for Flexible Non-Terminals in Dependency Tree-to-Tree Machine Translation *Nakazawa, Richardson, and Kurohashi*	Transition-Based Dependency Parsing with Heuristic Backtracking *Buckman, Ballesteros, and Dyer*	13:40
Who did What: A Large-Scale Person-Centered Cloze Dataset *Onishi, Wang, Bansal, Gimpel, and McAllester*	Why Neural Translations are the Right Length *Shi, Knight, and Yuret*	Word Ordering Without Syntax *Schmaltz, Rush, and Shieber*	13:55
Building compositional semantics and higher-order inference system for a wide-coverage Japanese CCG parser *Mineshima, Tanaka, Martínez-Gómez, Miyao, and Bekki*	Supervised Attentions for Neural Machine Translation *Mi, Wang, and Ittycheriah*	Morphological Segmentation Inside-Out *Cotterell, Kumar, and Schütze*	14:10
Learning to Generate Compositional Color Descriptions *Monroe, Goodman, and Potts*	Learning principled bilingual mappings of word embeddings while preserving monolingual invariance *Artetxe, Labaka, and Agirre*	Parsing as Language Modeling *Choe and Charniak*	14:25
A Decomposable Attention Model for Natural Language Inference *Parikh, Täckström, Das, and Uszkoreit*	Measuring the behavioral impact of machine translation quality improvements with A/B testing *Russell and Gillespie*	Human-in-the-Loop Parsing *He, Michael, Lewis, and Zettlemoyer*	14:40
Deep Reinforcement Learning for Mention-Ranking Coreference Models *Clark and Manning*	Creating a Large Benchmark for Open Information Extraction *Stanovsky and Dagan*	Unsupervised Timeline Generation for Wikipedia History Articles *Bauer and Teufel*	14:55
A Stacking Gated Neural Architecture for Implicit Discourse Relation Classification *Qin, Zhang, and Zhao*	Bilingually-constrained Synthetic Data for Implicit Discourse Relation Recognition *Wu, Chen, and Huang*	Encoding Temporal Information for Time-Aware Link Prediction *Jiang, Liu, Ge, Sha, Li, Chang, and Sui*	15:10

Session 8A: Short Paper Oral Session I

Salon FG

Tense Manages to Predict Implicative Behavior in Verbs
Ellie Pavlick and Chris Callison-Burch 13:40–13:55

Implicative verbs (e.g. "manage") entail their compliment clauses, while non-implicative verbs (e.g. "want") do not. For example, while "managing to solve the problem" entails "solving the problem", no such inference follows from "wanting to solve the problem." Differentiating between implicative and non-implicative verbs is therefore an essential component of natural language understanding, relevant to applications such as textual entailment and summarization. We present a simple method for predicting implicativeness which exploits known constraints on the tense of implicative verbs and their compliments. We show that this yields an effective, data-driven way of capturing this nuanced property in verbs.

Who did What: A Large-Scale Person-Centered Cloze Dataset
Takeshi Onishi, Hai Wang, Mohit Bansal, Kevin Gimpel, and David McAllester 13:55–14:10

We have constructed a new "Who-did-What" dataset of over 200,000 fill-in-the-gap (cloze) multiple choice reading comprehension problems constructed from the LDC English Gigaword newswire corpus. The WDW dataset has a variety of novel features. First, in contrast with the CNN and Daily Mail datasets (Hermann et al., 2015) we avoid using article summaries for question formation. Instead, each problem is formed from two independent articles — an article given as the passage to be read and a separate article on the same events used to form the question. Second, we avoid anonymization — each choice is a person named entity. Third, the problems have been filtered to remove a fraction that are easily solved by simple baselines, while remaining 84% solvable by humans. We report performance benchmarks of standard systems and propose the WDW dataset as a challenge task for the community.

Building compositional semantics and higher-order inference system for a wide-coverage Japanese CCG parser
Koji Mineshima, Ribeka Tanaka, Pascual Martínez-Gómez, Yusuke Miyao, and Daisuke Bekki 14:10–14:25

This paper presents a system that compositionally maps outputs of a wide-coverage Japanese CCG parser onto semantic representations and performs automated inference in higher-order logic. The system is evaluated on a textual entailment dataset. It is shown that the system solves inference problems that focus on a variety of complex linguistic phenomena, including those that are difficult to represent in the standard first-order logic.

Learning to Generate Compositional Color Descriptions
Will Monroe, Noah D. Goodman, and Christopher Potts 14:25–14:40

The production of color language is essential for grounded language generation. Color descriptions have many challenging properties: they can be vague, compositionally complex, and denotationally rich. We present an effective approach to generating color descriptions using recurrent neural networks and a Fourier-transformed color representation. Our model outperforms previous work on a conditional language modeling task over a large corpus of naturalistic color descriptions. In addition, probing the model's output reveals that it can accurately produce not only basic color terms but also descriptors with non-convex denotations ("greenish"), bare modifiers ("bright", "dull"), and compositional phrases ("faded teal") not seen in training.

A Decomposable Attention Model for Natural Language Inference
Ankur Parikh, Oscar Täckström, Dipanjan Das, and Jakob Uszkoreit 14:40–14:55

We propose a simple neural architecture for natural language inference. Our approach uses attention to decompose the problem into subproblems that can be solved separately, thus making it trivially parallelizable. On the Stanford Natural Language Inference (SNLI) dataset, we obtain state-of-the-art results with almost an order of magnitude fewer parameters than previous work and without relying on any word-order information. Adding intra-sentence attention that takes a minimum amount of order into account yields further improvements.

Deep Reinforcement Learning for Mention-Ranking Coreference Models
Kevin Clark and Christopher D. Manning 14:55–15:10

Coreference resolution systems are typically trained with heuristic loss functions that require careful tuning. In this paper we instead apply reinforcement learning to directly optimize a neural mention-ranking model for coreference evaluation metrics. We experiment with two approaches: the REINFORCE policy gradient algorithm and a reward-rescaled max-margin objective. We find the latter to be more effective, resulting in a significant improvement over the current state-of-the-art on the English and Chinese portions of the CoNLL 2012 Shared Task.

A Stacking Gated Neural Architecture for Implicit Discourse Relation Classification
Lianhui Qin, Zhisong Zhang, and Hai Zhao 15:10–15:25

Discourse parsing is considered as one of the most challenging natural language processing (NLP) tasks. Implicit discourse relation classification is the bottleneck for discourse parsing. Without the guide of explicit discourse connectives, the relation of sentence pairs are very hard to be inferred. This paper proposes a stacking neural network model to solve the classification problem in which a convolutional neural network (CNN) is utilized for sentence modeling and a collaborative gated neural network (CGNN) is proposed for feature transformation. Our evaluation and comparisons show that the proposed model outperforms previous state-of-the-art systems.

Session 8B: Short Paper Oral Session II

Salon J

Insertion Position Selection Model for Flexible Non-Terminals in Dependency Tree-to-Tree Machine Translation
Toshiaki Nakazawa, John Richardson, and Sadao Kurohashi 13:40–13:55

Dependency tree-to-tree translation models are powerful because they can naturally handle long range reorderings which is important for distant language pairs. The translation process is easy if it can be accomplished only by replacing non-terminals in translation rules with other rules. However it is sometimes necessary to adjoin translation rules. Flexible non-terminals have been proposed as a promising solution for this problem. A flexible non-terminal provides several insertion position candidates for the rules to be adjoined, but it increases the computational cost of decoding. In this paper we propose a neural network based insertion position selection model to reduce the computational cost by selecting the appropriate insertion positions. The experimental results show the proposed model can select the appropriate insertion position with a high accuracy. It reduces the decoding time and improves the translation quality owing to reduced search space.

Why Neural Translations are the Right Length
Xing Shi, Kevin Knight, and Deniz Yuret 13:55–14:10

We investigate how neural, encoder-decoder translation systems output target strings of appropriate lengths, finding that a collection of hidden units learns to explicitly implement this functionality.

Supervised Attentions for Neural Machine Translation
Haitao Mi, Zhiguo Wang, and Abe Ittycheriah 14:10–14:25

In this paper, we improve the attention or alignment accuracy of neural machine translation by utilizing the alignments of training sentence pairs. We simply compute the distance between the machine attentions and the "true" alignments, and minimize this cost in the training procedure. Our experiments on large-scale Chinese-to-English task show that our model improves both translation and alignment qualities significantly over the large-vocabulary neural machine translation system, and even beats a state-of-the-art traditional syntax-based system.

Learning principled bilingual mappings of word embeddings while preserving monolingual invariance
Mikel Artetxe, Gorka Labaka, and Eneko Agirre 14:25–14:40

Mapping word embeddings of different languages into a single space has multiple applications. In order to map from a source space into a target space, a common approach is to learn a linear mapping that minimizes the distances between equivalences listed in a bilingual dictionary. In this paper, we propose a framework that generalizes previous work, provides an efficient exact method to learn the optimal linear transformation and yields the best bilingual results in translation induction while preserving monolingual performance in an analogy task.

Measuring the behavioral impact of machine translation quality improvements with A/B testing
Ben Russell and Duncan Gillespie 14:40–14:55

In this paper we discuss a process for quantifying the behavioral impact of a domain-customized machine translation system deployed on a large-scale e-commerce platform. We discuss several machine translation systems that we trained using aligned text from product listing descriptions written in multiple languages. We document the quality improvements of these systems as measured through automated quality measures and crowdsourced human quality assessments. We then measure the effect of these quality improvements on user behavior using an automated A/B testing framework. Through testing we observed an increase in key e-commerce metrics, including a significant increase in purchases.

Creating a Large Benchmark for Open Information Extraction
Gabriel Stanovsky and Ido Dagan 14:55–15:10

Open information extraction (Open IE) was presented as an unrestricted variant of traditional information extraction. It has been gaining substantial attention, manifested by a large number of automatic Open IE extractors and downstream applications. In spite of this broad attention, the Open IE task

definition has been lacking – there are no formal guidelines and no large scale gold standard annotation. Subsequently, the various implementations of Open IE resorted to small scale post-hoc evaluations, inhibiting an objective and reproducible cross-system comparison. In this work, we develop a methodology that leverages the recent QA-SRL annotation to create a first independent and large scale Open IE annotation, and use it to automatically compare the most prominent Open IE systems.

Bilingually-constrained Synthetic Data for Implicit Discourse Relation Recognition
Changxing Wu, Yidong Chen, and Yanzhou Huang 15:10–15:25

To alleviate the shortage of labeled data, we propose to use bilingually-constrained synthetic implicit data for implicit discourse relation recognition. These data are extracted from a bilingual sentence-aligned corpus according to the implicit/explicit mismatch between different languages. Incorporating these data via a multi-task neural network model achieves significant improvements over baselines, on both the English PDTB and Chinese CDTB data sets.

Session 8C: Short Paper Oral Session III

Room 616 AB

Transition-Based Dependency Parsing with Heuristic Backtracking

Jacob Buckman, Miguel Ballesteros, and Chris Dyer 13:40–13:55

We introduce a novel approach to the decoding problem in transition-based parsing: heuristic backtracking. This algorithm uses a series of partial parses on the sentence to locate the best candidate parse, using confidence estimates of transition decisions as a heuristic to guide the starting points of the search. This allows us to achieve a parse accuracy comparable to beam search, despite using fewer transitions. When used to augment a Stack-LSTM transition-based parser, the parser shows an unlabeled attachment score of up to 93.30% for English and 87.61% for Chinese.

Word Ordering Without Syntax

Allen Schmaltz, Alexander M. Rush, and Stuart Shieber 13:55–14:10

Recent work on word ordering has argued that syntactic structure is important, or even required, for effectively recovering the order of a sentence. We find that, in fact, an n-gram language model with a simple heuristic gives strong results on this task. Furthermore, we show that a long short-term memory (LSTM) language model is even more effective at recovering order, with our basic model outperforming a state-of-the-art syntactic model by 11.5 BLEU points. Additional data and larger beams yield further gains, at the expense of training and search time.

Morphological Segmentation Inside-Out

Ryan Cotterell, Arun Kumar, and Hinrich Schütze 14:10–14:25

Morphological segmentation has traditionally been modeled with non-hierarchical models, which yield flat segmentations as output. In many cases, however, proper morphological analysis requires hierarchical structure—especially in the case of derivational morphology. In this work, we introduce a discriminative joint model of morphological segmentation along with the orthographic changes that occur during word formation. To the best of our knowledge, this is the first attempt to approach discriminative segmentation with a context-free model. Additionally, we release an annotated treebank of 7454 English words with constituency parses, encouraging future research in this area.

Parsing as Language Modeling

Do Kook Choe and Eugene Charniak 14:25–14:40

We recast syntactic parsing as a language modeling problem and use recent advances in neural network language modeling to achieve a new state of the art for constituency Penn Treebank parsing — 93.8 F1 on section 23, using 2-21 as training, 24 as development, plus tri-training. When trees are converted to Stanford dependencies, UAS and LAS are 95.9% and 94.1%.

Human-in-the-Loop Parsing

Luheng He, Julian Michael, Mike Lewis, and Luke Zettlemoyer 14:40–14:55

This paper demonstrates that it is possible for a parser to improve its performance with a human in the loop, by posing simple questions to non-experts. For example, given the first sentence of this abstract, if the parser is uncertain about the subject of the verb “pose,” it could generate the question What would pose something? with candidate answers this paper and a parser. Any fluent speaker can answer this question, and the correct answer resolves the original uncertainty. We apply the approach to a CCG parser, converting uncertain attachment decisions into natural language questions about the arguments of verbs. Experiments show that crowd workers can answer these questions quickly, accurately and cheaply. Our human-in-the-loop parser improves on the state of the art with less than 2 questions per sentence on average, with a gain of 1.7 F1 on the 10% of sentences whose parses are changed.

Unsupervised Timeline Generation for Wikipedia History Articles

Sandro Bauer and Simone Teufel 14:55–15:10

This paper presents a generic approach to content selection for creating timelines from individual history articles for which no external information about the same topic is available. This scenario is in contrast to existing works on timeline generation, which require the presence of a large corpus of news articles. To identify salient events in a given history article, we exploit lexical cues about the article’s subject area, as well as time expressions that are syntactically attached to an event word. We also test

different methods of ensuring timeline coverage of the entire historical time span described. Our best-performing method outperforms a new unsupervised baseline and an improved version of an existing supervised approach. We see our work as a step towards more semantically motivated approaches to single-document summarisation.

Encoding Temporal Information for Time-Aware Link Prediction
Tingsong Jiang, Tianyu Liu, Tao Ge, Lei Sha, Sujian Li, Baobao Chang, and Zhifang Sui
15:10–15:25

Most existing knowledge base (KB) embedding methods solely learn from time-unknown fact triples but neglect the temporal information in the knowledge base. In this paper, we propose a novel time-aware KB embedding approach taking advantage of the happening time of facts. Specifically, we use temporal order constraints to model transformation between time-sensitive relations and enforce the embeddings to be temporally consistent and more accurate. We empirically evaluate our approach in two tasks of link prediction and triple classification. Experimental results show that our method outperforms other baselines on the two tasks consistently.

Session P4: Best Paper Plenary Session

Salon FG *Chair: Kevin Duh and Xavier Carreras*

Improving Information Extraction by Acquiring External Evidence with Reinforcement Learning
Karthik Narasimhan, Adam Yala, and Regina Barzilay 15:55–16:20

Most successful information extraction systems operate with access to a large collection of documents. In this work, we explore the task of acquiring and incorporating external evidence to improve extraction accuracy in domains where the amount of training data is scarce. This process entails issuing search queries, extraction from new sources and reconciliation of extracted values, which are repeated until sufficient evidence is collected. We approach the problem using a reinforcement learning framework where our model learns to select optimal actions based on contextual information. We employ a deep Q-network, trained to optimize a reward function that reflects extraction accuracy while penalizing extra effort. Our experiments on two databases – of shooting incidents, and food adulteration cases – demonstrate that our system significantly outperforms traditional extractors and a competitive meta-classifier baseline.

Global Neural CCG Parsing with Optimality Guarantees
Kenton Lee, Mike Lewis, and Luke Zettlemoyer 16:20–16:45

We introduce the first global recursive neural parsing model with optimality guarantees during decoding. To support global features, we give up dynamic programs and instead search directly in the space of all possible subtrees. Although this space is exponentially large in the sentence length, we show it is possible to learn an efficient A* parser. We augment existing parsing models which have informative bounds on the outside score, with a global model that has loose bounds but only needs to model non-local phenomena. The global model is trained with a new objective that encourages the parser to explore a tiny fraction of the search space. The approach is applied to CCG parsing, improving state-of-the-art accuracy by 0.4 F1. The parser finds the optimal parse for 99.9% of held-out sentences, exploring on average only 190 subtrees.

Learning a Lexicon and Translation Model from Phoneme Lattices
Oliver Adams, Graham Neubig, Trevor Cohn, Steven Bird, Quoc Truong Do, and Satoshi Nakamura 16:45–17:10

Language documentation begins by gathering speech. Manual or automatic transcription at the word level is typically not possible because of the absence of an orthography or prior lexicon, and though manual phonemic transcription is possible, it is very slow. On the other hand, translations of the minority language into a major language are more easily acquired. We propose a method to harness such translations to improve automatic phoneme recognition. The method assumes no prior lexicon or translation model, instead learning them from phoneme lattices and translations of the speech being transcribed. Experiments demonstrate phoneme error rate improvements against two baselines and the model's ability to learn useful bilingual lexical entries.

SQuAD: 100,000+ Questions for Machine Comprehension of Text
Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang 17:10–17:35

We present the Stanford Question Answering Dataset (SQuAD), a new reading comprehension dataset consisting of 100,000+ questions posed by crowdworkers on a set of Wikipedia articles, where the answer to each question is a segment of text from the corresponding reading passage. We analyze the dataset to understand the types of reasoning required to answer the questions, leaning heavily on dependency and constituency trees. We build a strong logistic regression model, which achieves an F1 score of 51.0%, a significant improvement over a simple baseline (20%). However, human performance (86.8%) is much higher, indicating that the dataset presents a good challenge problem for future research. The dataset is freely available at https://stanford-qa.com.

Index


Google

Research & Academic Relations

We are committed to supporting and building strong relationships with the academic community in a variety of different ways, including:

- Facebook Fellowship
- Emerging Scholars Award
- Research Collaborations
- Conference Sponsorships
- Faculty Sabbaticals, Invited Talks, and Visiting Scholars
- Internships, Full Time Positions, Post Docs

research.facebook.com

facebook

Microsoft

How To Get Here From the Austin-Bergstrom International Airport

DRIVING DIRECTIONS

Take Hwy 71 West approximately 5.5 miles to I-35. Merge onto I-35 North and take exit 234C. Turn left at the 1st light (6th St). Travel west on 6th St and turn left onto Neches St. The Hilton Austin is on the corner of 5th and Neches.

Distance from Hotel: 7 mi.
Drive Time: 15 min.

TRANSPORTATION TO AND FROM AIRPORT:

Type	Typical Minimum Charge
Bus Service	1.75 USD
Limousine	Fees vary. Contact hotel.
Super Shuttle	16.00 USD
Rental Car	Fees vary.
Taxi	30.00 USD

Transportation from Austin-Bergstrom International Airport to and from the Hilton is available with SuperShuttle “Shared Ride" or ExecuCar “Private Sedan/SUV". Make reservations at www.SuperShuttle.com or www.ExecuCar.com use code HILTN for a 5% discount off published online fare.

Food Guide

Barbecue	Mexican	American	Italian	Seafood
Lamberts (W.2nd/Guadalupe) **Iron Works** (Red River/Cesar Chavez) **Stubbs** (Red River/E.8th) **Franklin** (E.11th/Branch) **La Barbecue** (E.Cesar Chavez) **Freedman's** (W. 24th) **Terry Blacks** (Barton Springs Rd) **Micklethwait** (E.11th/Rosewood) **The Salt Lick** (Round Rock/Driftwood)	**La Condesa** (W.2nd/Guadalupe) **El Naranjo** (Rainey St.) **Cantina Laredo** (W.3rd/Colorado) **Manuels** (Congress/W.3rd) **Iron Cactus** (E.6th St) **Pelons** (8th/Red River) **Benjis Cantina** (W.6th/West Ave) **Uncle Julios** (Brazos/E.3rd) **El Sol Y La Luna** (6th/Red River) **Matts El Rancho** (S. Lamar)	**Moonshine** (4TH/Red River) **Parkside** (E.6th /San Jacinto) **Searsucker** (Colorado/W.5th) **Fixe** (W.5th/Nueces) **Max's Wine Dive** (San Jacinto/E.3rd) **Lonesome Dove** **(5th/Colorodo)** **Second Bar + Kitchen** (Congress/W.2nd) **The Bonneville** (Cesar Chavez/Colorado) **No Va** (Rainey St.) **24 Diner** (S.Lamar/W. 6th)	**Italic** (W.6th/Colorado) **Botticellis** (S.Congress/Gibson) **La Traviata** (Congress/W. 4th) **Taverna** (W.2nd/Lavaca) **Quattro Gatti** (Congress/ W.9th) **WinfloOsteria** (W.6th/Pressler) **Carmelos** (E.5t) **Cipollina** (W.Lynn/W. 13th) **Vespaio** (S. Congress)	**Finn and Porter** (Hilton Austin) **Trulucks** (Colorado/E.4th) **Perla's** (S. Congress/Gibson) **Eddie V's** (E. 5th/Trinity) **Surf and Turf Po Boy** (Lavaca/E. 4th) **Clarks** (W. 6th/Blanco) **Trio** (San Jacinto) **Capital Grille** (W. 4th/Colorado) **Dock and Roll** (S.1st)
Steakhouse	**Asian**	**Sushi**	**French**	**Pizza**
Finn and Porter (Hilton Austin) **Eddie V's** (E. 5th/San Jacinto) **Vince Young's** (2nd/San Jacinto) **Capital Grille** (W. 4th/Colorado) **Perry's** (W.7th/Colorado) **Sullivan's** (Colorado/W.3rd) **III Forks** (C. Chavez/Lavaca) **Austin Land & Cattle** (N.Lamar)	**Qui** (E. 6th) **Imperia** (San Antonio/5th) **Kyoten** **(E. 6th)** **Mai Thai** (San Jacinto/E.3rd) **Sway** (S.1ST) **Elizabeth St. Café** (S.1st/W.Mary) **Koriente** (E.7th/Sabine) **Daruma Ramen** (E.6th)	**Finn and Porter** (Hilton Austin) **Bar-Chi** (Colorado/W.2nd) **Uchi** (S.Lamar) **Kome** (Airport Blvd) **Sushi Zushi** (W.5th) **Piranha Sushi** (San Jacinto/E.3rd) **Maiko** (W. 6th/Guadalupe) **RA** (4th/Colorado)	**Arro** (W.6th/Nueces) **Justine's Brasserie** (E. 5th) **LaV** (E. 7th) **Chez Nous** (5th/Neches) **Peche** (W.4th) **Blue Dahlia Bistro** (E.11th)	**Backspace** (San Jacinto/E 6th) **DueForni** (E.6th/Congress) **Hoboken Pie** (Red River/E.8th) **Bufalina** (E. Cesar Chavez) **Home Slice** (S. Congress) **East Side Pies** (Rosewood) **Via 313** (Rainey St)
Burgers/Sandwiches	**Vegetarian**	**Indian/Misc.**	**Breakfast**	**Brunch**
Eureka (6th/ Brazos) **Hopdoddy** (S.Congress) **Easy Tiger** (E.6th/Sabine) **Waller Creek Pub** (6th/ Sabine) **Casino El Camino** (E.6th/Red River) **P Terry's** (6th/Congress) **Frank** (Colorado/W.4th)	**Gardner** (E. 6th) **Koriente** (E.7th/Sabine) **Arlos** (Red River) **Counter Culture Café** (E. C. Chavez/Clara) **Bouldin Creek Café** (S.1st/W.Mary) **Mr. Natural** (E. C. Chavez/Chicon) **The Vegan Yacht** (Spiderhouse)	**Nasha Indian/Tex Mex** (E. 7th) **G'RajMahal Indian** (Rainey St) **Malaga Tapas** (2nd/ San Antonio) **Clay Pit** (15th/Guadalupe) **Russian House** (E. 5th) **Buenos Aires Argentine** (E. 6th/Waller) **Bangers** (Rainey St)	**Liberty Tavern** (Hilton Austin) **24 Diner** (S.Lamar/W.6th) **Annies Cafe** (Congress/E.4th) **1886 Café** (The Driskill) **Counter Café** (E.6th) **Sawyer and Co.** (E. Cesar Chavez) **South Congress Café** (South Congress) **Snack Bar** (S.Congress)	**Moonshine** (Red River/E.4th) **El Naranjo** (Rainey St) **Austin Ale House** (W. 6th) **Stubbs Gospel Brunch** (Red River) **No Va** (Rainey) **Swifts Attic** (Congress Avenue) **Taverna** (2nd/Lavaca) **Searsucker** (5th/Colorado)
Cafes	**Beer/Wine/Liquor**	**Grocery/Pharmacy**	**Tour Info/Pickup**	
Java Jive (Hilton Austin) **Hounds Tooth** (E.4th/Congress) **Café Medici** (Congress)	**All American Liquors** (E 5th/Brazos) **Twin Liquors** (Red River/E.7th) **Austin Wine Merchant** (W.6th)	**CVS Pharmacy** (Congress/W.5th) **Royal Blue Grocery** (3rd/Brazos) **Whole Foods Maket** (W.6th/Lamar)	**Austin Visitor Center** (512)-478-0098 (Red River/E.4th)	**The Hilton Austin** **500 E. 4th St** **Austin, Texas** **(512)-482-8000**

Area Map


TEXAS STATE CAPITOL
Starbucks
Mayhem
Subway
The Clay Pit
CAMPUS "THE DRAG"
DoubleTree Hotel
Texas Chili Parlor
15TH
14TH
13TH
12TH
11TH
10TH
9TH
8TH
7TH
6TH
5TH
4TH
3RD
2ND
Waterloo Park
Brick Oven
Serrano's
La Quinta Inn
Franklins
SABINE ST
AVIS Rent a Car
SHERATON
Hertz Rent a Car
CIVIC DISTRICT
COUNTY COURT HOUSE
GOVERNOR'S MANSION
WOOLDRIDGE SQUARE
AUSTIN HISTORY CENTER LIBRARY
NUECES
SAN ANTONIO
GUADALUPE
LAVACA
COLORADO
BRAZOS
SAN JACINTO
TRINITY
NECHES
RED RIVER
CONGRESS AVENUE
Quattro Gatti Quiznos
AUSTIN CLUB
ST. MARYS CATHEDRAL
First Baptist Church
Mohawk Cheer Up Charlies Arlos
Chipotle Subway
Pelons 508 Tequila Bar
Stubb's BBQ
Paramount CU29 Roaring Fork
Townsend Austin Museum Cont. Art Perry's
InterContinental
Omni Hotel
ST. DAVIDS EPISCOPAL CHURCH
Valhalla Hoboken Pie
Elysium Side Bar SideWinder Empire
616 Ranch Rattle Inn Slate Rio
New Movement Comedy Glorias
Athenian Grill Starbuck's
7-Eleven Royal Blue Capital One Due Forni
FireHouse Voodoo Friends Bar BD Riley's Eureka
Alamo Drafthouse Iron Cactus Shakespeares Museum of Weird
Tellers Vulcan Gas Roppollos Jackalope
Twin Liquors
Barbarella Koriente Daruma Ramen El Sol y La Luna
Waller Creek The Gatsby Gulf Gas Station
MARKET DISTRICT
ENTERTAINMENT DISTRICT
East Side
Frio Hofbrau Arro Key Bar Waltons Wu Chow
RIO BENJIS WHOLE FOODS 24 DINER WWFLO OSTERIA Z' TEJAS
Maiko The Belmont Austin Ale House
Chase Headliner Club
Ruths Chris Steak House
CVS pharmacy
P. Terrys Wild About Music Shop Keepers Jimmy John's
Buffalo Billards Barcelona Chuggin Monkey
Maggie Maes Parkside Backspace Old Pecan Stella San Jac
Old School The Library Chez Nous Pete's Plano
Casino El Camino Esthers Follies Carmelo's
Jim Jim's Water Ice Austin Segway Tours
Easy Tiger
Hilton Garden Inn
Juan Pelota Kung Fu Mellow Johnny's Bike Shop
Republic Square Park
Rebels Saloon Madison Republic Lake
Gourdough's Halycon Peche Fado Irish Pub Trulucks
Lonesome Dove Chinatown searsucker Franks
The Handlebar Houndstooth Taco Shack BlueRibbon
Good Life Barbershop Blenders & Bowls
Russian House Eddie V's Champions
O'Henry Museum Brush Square
Firm&Porter Liberty Tavern Java Jive
Visitors Bureau
Conv. Center Parking
Gas Pipe
WAREHOUSE DISTRICT
Front Door
Prelog's
Capital Grill RA Sushi Gingerman Sullivan's
Manuel's La Traviata
FedEx Office Annie's Elephant Room Counter 3 57 SwiftsAttic
Royal Blue Grocery Uncle Julios
VinceYoung Steakhouse Red Room
Moonshine
St. Bernard
Violet Crown Cinema La Condesa Malaga Austin MacWorks
Trace
Austin City Limits
Royal Blue Grocery Bar Chr Cantina Laredo Taverna Cru Jo's Coffee
Café Medici Congress 2nd Bar+Kitchen Berrys Frozen Yogurt
W Austin Burger Bar Corner Bar Osterio Pronto
Hyatt Place
Cedar Door Café Crepe
Fogo de Chao Max's Wine Dive Mai Thai PFChang's Fleming's Piranha's Crave
AUSTIN CONVENTION CENTER
PALM PARK
SHOPPING DISTRICT
WILLIE NELSON BLVD (2ND)
Austin Java Bonneville
Austin Rocks
Ill Forks Mama Fu's
Wells Fargo Murphy's Deli
Gus's Fried Chicken Michetadas Brass House
Iron Works BBQ
CESAR CHAVEZ (1ST)
CONGRESS "BAT" BRIDGE
Radisson
Dine
Four Seasons
Alta's Cafe
Kayak & Bike Rentals
DRISKILL
Container Bar
Bar 96 Blackheart Clive
El Naranjo Nu Va Bangers Craft Pride VIA 313
RAINEY
LADY BIRD LAKE
B. Springs Picnic
palmer
Long Center
CAPITOL CRUISES
HYATT
To Zilker Park
BARTON SPRINGS RD
EMBASSY SUITES
SOCO
Downtown Austin

EMNLP 2016 gratefully acknowledges the following sponsors for their support:

Platinum

Gold

Silver

Bronze

Student Volunteer Sponsor