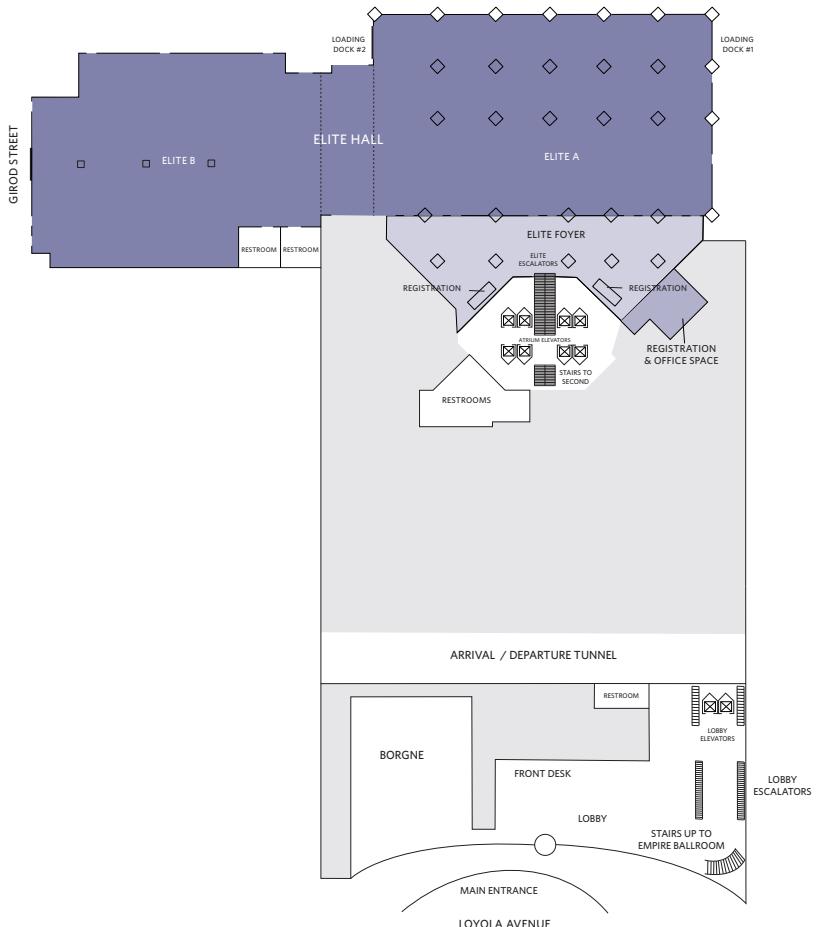


NAACL HLT

New Orleans, Louisiana | June 1-June 6, 2018

Conference Handbook

The 16th Annual Conference of The North American Chapter of The Association For Computational Linguistics: Human Language Technologies



Cover design by Ying Lin

Cover photo is a derivative of original work by Peyton Stanton available at <https://www.flickr.com/photos/peytonstanton/35496401710/>
<https://creativecommons.org/licenses/by-nc/2.0/>



Handbook assembled by Jonathan May

Printing by Omnipress of Madison, Wisconsin

Contents

Table of Contents	i
1 Conference Information	1
Message from the General Chair	1
Message from the Program Co-Chairs	3
Message from the Industry Track Co-Chairs	5
Organizing Committee	7
Program Committee	9
Outstanding Reviewers	11
Ad-hoc Reviewers	12
Meal Info	13
2 Workshops and Tutorials: Friday, June 1	15
WiNLP: Widening NLP	16
Message from the Tutorial Co-Chairs	20
T1: Modelling Natural Language, Programs, and their Intersection	21
T2: Deep Learning Approaches to Text Production	22
T3: Scalable Construction and Reasoning of Massive Knowledge Bases	23
T4: The Interplay Between Lexical Resources and Natural Language Processing	25
T5: Socially Responsible NLP	26
T6: Deep Learning for Conversational AI	27
Welcome Reception: Friday, June 1, 18:00	29
Welcome Reception	29
3 Main Conference: Saturday, June 2	31
Keynote Address: Charles Yang (sponsored by Toutiao AI Lab)	32
Session 1	33
Oral: Information Extraction 1	34
Oral: Phonology, Morphology and Word Segmentation 1	34
Oral: Speech 1	35
Industry: Machine Learning – Classification	36

Session 2	37
Oral: Machine Learning 1	38
Oral: Information Extraction 2	38
Oral: Machine Translation 1	39
Industry: Dialog	40
Posters and Demos: Sessions 1/2	41
Posters: Discourse and Pragmatics 1	41
Posters: Generation 1	43
Posters: NLP Applications 1	46
Demos	49
Recruitment Lunch	51
Industry Keynote Address: Mari Ostendorf	52
Session 3	53
Oral: Machine Learning 2	54
Oral: Social Media and Computational Social Science 1	55
Oral: Vision, Robotics and Other Grounding 1	55
Industry Panel: Careers in the Industry	56
Posters: Semantics 1	57
Posters: Sentiment Analysis 1	62
Demos	65
Session 4	66
Oral: NLP Applications 2	67
Oral: Question Answering 1	68
Oral: SRW Highlight Papers	69
Industry: Machine Translation	70
4 Main Conference: Sunday, June 3	73
Keynote Address: Kevin Knight (sponsored by Google)	74
Session 5	75
Oral: Cognitive Modeling and Psycholinguistics 1	76
Oral: Summarization 1	76
Oral: Semantics 2	77
Industry: Machine Learning – Domain Adaptation	78
Session 6	79
Oral: Machine Learning 3	80
Oral: Social Media and Computational Social Science 2	80
Oral: Vision, Robotics and Other Grounding 2	81
Industry: Named Entity and Language Generation	82
Posters and Demos: Sessions 5/6	83
Posters: Information Extraction 3	83
Posters: Tagging, Chunking, Syntax and Parsing 1	88
Demos	92
Industry Keynote Address: Daniel Marcu	93
Session 7	94
Oral: Text Mining 1	95
Oral: Semantics 3	96
Oral: Tagging, Chunking, Syntax and Parsing 2	97
Industry Panel: Ethics in NLP	98
Posters: Machine Learning 4	99
Posters: Machine Translation 2	102
Posters: Phonology, Morphology and Word Segmentation 2	106
Demos	108
Session 8	109
Test-of-Time Paper Session (in memory of Aravind Joshi)	110

Social Event: Sunday, June 3, 18:45	111
Social Event	111
5 Main Conference: Monday, June 4	113
Keynote Address: Dilek Hakkani-Tür (sponsored by Bloomberg)	114
Session 9	115
Information Extraction 4	116
Semantics 4	116
Generation 2	117
Session 10	119
Oral: Dialogue and Interactive Systems 1	120
Oral: Information Extraction 5	120
Oral: Generation 3	121
Industry: Machine Learning - Prediction	122
Posters and Demos: Sessions 9/10	124
Posters: Question Answering 2	124
Posters: Social Media and Computational Social Science 3	126
Posters: Summarization 2	128
Demos	133
Session 11	134
Oral: Sentiment Analysis 2	135
Oral: Discourse and Pragmatics 2	136
Oral: Tagging, Chunking, Syntax and Parsing 3	137
Posters: Cognitive Modeling and Psycholinguistics 2	139
Posters: Dialogue and Interactive Systems 2	140
Posters: Text Mining 2	142
Posters: Speech 2	143
Posters: Vision, Robotics and Other Grounding 3	144
Demos	147
Session 12	149
Outstanding Paper Session (sponsored by Amazon)	150
6 Workshops and Collocated Events: Tuesday–Wednesday, June 5–6	151
SemEval: 12th International Workshop on Semantic Evaluation	152
*SEM: Seventh Joint Conference On Lexical And Computational Semantics	165
GenDeep: Generalization in the Age of Deep Learning	168
BEA: 13th Workshop on Innovative Use of NLP for Building Educational Applications	169
Ethics-NLP: Second ACL Workshop on Ethics in Natural Language Processing	173
CLPsych: Fifth Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality	174
SemBEaR: Computational Semantics Beyond Events and Roles	176
Style-Var: 2nd Workshop on Stylistic Variation	177
Story-NLP: Workshop on Storytelling	178
SCLeM: Subword and Character LEvel Models in NLP	179
TextGraphs: Twelfth Workshop on Graph-Based Methods for Natural Language Processing	180
FigLang: Workshop on Figurative Language Processing	181
PEOPLES: Second Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media	183
CRAC: Workshop on Computational Models of Reference, Anaphora and Coreference	185
SpLU: First International Workshop on Spatial Language Understanding	186
7 Anti-Harassment Policy	187
8 Local Information	189

9 On-Site Childcare	193
Author Index	195
Sponsorship	218

Conference Information

Message from the General Chair

Welcome to New Orleans and to NAACL HLT 2018 – the biggest NAACL to date. Natural Language Processing and Computational Linguistics is constantly growing and changing with a constant flow of new methods and topics. Every year also sees an even more exciting and diverse research community, with a steadily increasing number researchers, companies both large and small, and a vibrant community of practitioners and students who are excited at the prospect of taking on the newest challenges of the discipline. This year's NAACL HLT conference reflects what an exciting time this is for our field, and highlights the vibrancy and vitality of our community.

I feel extremely lucky to be able to work with a fantastic program committee, especially the two extremely dedicated, creative and resourceful program chairs: Amanda Stent and Heng Ji. Their innovations include a new review form, intended to elicit higher quality reviews, the opportunity for authors to review the reviewers, the Test-of-Time awards, and a program where poster and demo sessions run consistently in parallel to the oral sessions, in order to allow the conference to reflect the ever increasing diversity of research topics and the corresponding volume of accepted papers. I am especially excited about the new Test-of-Time papers award session, and hope to see this new innovation become a regular part of ACL conferences.

We have named the Test-of-Time award in memory of Aravind Joshi, who left us this year, after having a huge lifetime impact on our community. We will always remember him for his gentle conversational style, sharp focus, interest in linguistic, computational and mathematical properties of language, and his lifetime commitment to mentoring women in NLP. I feel extremely lucky to have been one of his Ph.D. students.

This year we also introduced an industrial track, with the aim of featuring papers that focus on scalable, interpretable, reliable and customer facing methods for industrial applications of Natural Language Processing. The idea of having such a track was proposed by Yunyao Li who strongly advocated for it: this proposal was then discussed and approved by the NAACL board. After that, it was all go, with an incredible amount of work to promote and organize it by the industrial track chairs: Jennifer Chu-Carroll, Yunyao Li and Srinivas Bangalore.

The overall program looks amazing and reflects the cooperative way that everyone on the committee worked together. What a team! I am so grateful for getting to be a part of this community of people, and I really appreciate the enthusiasm and attention to detail reflected in their hard work: Amanda Stent and Heng Ji (program chairs); Jennifer Chu-Carroll, Yunyao Li and Srinivas Bangalore (industrial track chairs); Ying Lin (website chair); Marie Metteer and Jason Williams (workshop co-chairs); Mohit Bansal and Rebecca Passonneau (tutorial co-chairs); Yang Liu, Tim Paek, and Manasi Patwardhan (demo co-chairs); Chris Callison-Burch and Beth Hockey (Family-Friendly Program Co-Chairs) Stephanie Lukin and Meg Mitchell (publication

Message from the General Chair

co-chairs); Jonathan May (handbook chair); Silvio Ricardo Cordeiro, Shereen Oraby, Umashanthi Pavalanathan, and Kyeongmin Rim (student cochairs) along with Swapna Somasundaran and Sam Bowman (Faculty Advisors) for the student research workshop; Lena Reed (student volunteer coordinator); Kristy Hollingshead, Kristen Johnson, and Parisa Kordjamshidi (local sponsorships and exhibits cochairs); Yonatan Bisk and Wei Xu (publicity and social media chairs); David Yarowsky and Joel Tetreault (treasurers) and Alexis Palmer and Jason Baldridge (the NAACL international Sponsorship Team). Also thanks to Rich at SoftConf for his speedy response to questions and his willingness to help us innovate with our new review form. And thanks to Julia Hockenmaier and the whole NAACL Executive Board for always being willing to consult on any issue.

The program highlights three keynote speakers in the main track: Dilek Hakkani-Tür, Kevin Knight, and Charles Yang. We also have two keynote speakers in the industry track: Mari Ostendorf and Daniel Marcu. These talks promising to be interesting across a range of issues from language acquisition in children to the commercial possibilities of conversational agents. The industry track will also feature two panels, one on careers in industry (as compared to academia) and the other on ethics in NLP. The program also includes six tutorials featuring topics of current interest and sources of innovation in the field. We have sixteen workshops plus the student research workshop: some of these workshops have become events in themselves with many of them repeated each year. We will also have plenary sessions for the outstanding paper awards and the new Test-Of-Time papers award session.

Any event of this scale can only happen with the hard work of a wonderful group of people. I especially want to thank the NAACL board for being willing to consult on a range of different issues and Priscilla Rasmussen for taking care of all the millions of details that need to be looked after every single day to make sure the logistical aspects of the conference come together. I want to especially thank Priscilla for her hard work and creativity organizing our social event: we first will go to Mardi Gras World to see the world of wonders created each year for the Mardi Gras. From there we go to the river, to the dockside River City Plaza and River City Ballroom for New Orleans' famous cuisine and libations and dancing to live Zydeco, funk, soul and R&B.

ACL has been working for several years to increase diversity at our conferences and in our community. So, taking inspiration from ACL 2017, we aimed to make NAACL family friendly, by providing childcare at the conference, and encouraging people to bring their families to the social events and breakfasts. Diversity can also be a consequence of the support for students to attend the conference that we receive from the NSF, CRA-CCC and CRA-W: this subsidizes student travel to the student research workshop as well as their registration and ACL memberships. When combined with the support we are able to give to our student volunteers, we aim to make it possible for students from all over the world to come to the conference and be part of our community. We also decided, in consultation with the NAACL board, to provide subsidies to the Widening NLP workshop, which is only being held for the second time at this year's NAACL (last year called the Women in NLP workshop). These subsidies enable participation from students and young researchers from developing countries to attend the conference.

I am grateful to our sponsors for their generous contributions, which add so much to what we can do at the conference. Our Diamond sponsors are Bloomberg, Google, and Toutiao AI Lab. The Platinum sponsor is Amazon. The Gold Sponsors are Ebay, Grammarly, IBM Research, KPMG, Oracle, Poly AI, Tulane University, Capital One and Two Sigma. The Silver sponsors are Nuance and Facebook, and the Bronze sponsors are iMerit and USC/ISI.

Finally, there are many more people who through their hard work and dedication have contributed to make this conference a success: the area chairs, workshop organizers, tutorial presenters, student mentors, and reviewers. And of course you all, the attendees without whom there would be no conference: you are the life and spirit of the conference and the NAACL community. I hope you all have a fun and exciting time at NAACL HLT 2018!

NAACL HLT 2018 General Chair
Marilyn Walker, University of California Santa Cruz

Message from the Program Co-Chairs

We welcome you to New Orleans for the 16th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL HLT 2018)! We had three primary goals for NAACL HLT 2018: construct a great program; manage the integrity and quality of the publication process; and ensure broad participation.

Construct a great program: NAACL HLT 2018 does have a great program, thanks to all of you! We will have three exciting keynotes, from Charles Yang, Kevin Knight and Dilek Hakkani-Tür. 331 research track papers (205 long, 125 short), accepted following peer review, will be presented¹. Four of these papers have been identified as outstanding papers, and one will be named best paper. We will also feature a “Test of Time” session with retrospectives (from the authors) on three influential papers from ACL venues. We thank the committees who nominated and voted on these paper awards.

The main program at NAACL HLT 2018 also includes 16 TACL paper presentations, 20 demos, a student research workshop and an industry track. Keynotes from both the research and industry tracks are plenary. In a change from previous years of NAACL HLT, and motivated by EMNLP 2017, poster and oral presentation sessions will be held in parallel during the day. All posters are grouped thematically (including posters from the industry track and student research workshop and demos) and assigned to poster sessions so as to not be against oral presentation sessions with the same theme.

Manage the integrity and quality of the publication process: To manage load, we decided that each area chair should be responsible for no more than 30 submissions and that reviewers should be responsible for reviewing no more than 3 submissions. To help reviewers, we and the ACL program co-chairs constructed a more structured review form, with questions related to the new ACL guidelines on publication and reviewing, as well as to contribution types, experimental methods (thank you, Bonnie Webber!), software and handling of data.

We recruited an excellent group of 72 area chairs; we thank them for their leadership, and for nominating and voting on outstanding papers, outstanding reviewers and test of time papers. 1372 individuals reviewed papers for the conference (as program committee members, ad hoc reviewers or secondary reviewers); all but 49 reviewers had no more than 3 submissions to review overall, and the 49 reviewers who took on a heavier load did so voluntarily. We thank all our reviewers, especially the ad-hoc reviewers who provided last minute reviews and the outstanding reviewers identified by the area chairs.

Submissions were assigned to area chairs and reviewers using a combination of area chair expertise, Toronto Paper Matching System (TPMS) scores and reviewer bids. Both long and short paper submissions received 3 reviews each. Long paper authors had an opportunity to respond to reviews. First-round accept/reject suggestions were made by area chairs working in small groups of 2-3 and discussing with the reviewers as necessary; final decisions were made by the program chairs. Where the reviews disagreed or there was considerable discussion, one area chair wrote a short meta review that was shared with the authors.

This year, if the authors of a NAACL HLT 2018 submission and the author of a review for that submission both consented, then we will include the review in a review corpus to be released jointly with the program chairs of ACL, Iryna Gurevich and Yusuke Miyao. We also asked authors of accepted papers to upload the source code for their papers. Both of these corpora will be released in the coming months.

The health of our field as a science is dependent on a scalable peer review process, which in turn depends on (a) conscientious effort from a broad pool of expert reviewers, and (b) tools, processes and policies that can scaffold and enable reviewing. As a field we are at a breaking

¹We received 1122 research track submissions (664 long, 458 short). 33 were rejected without review and 85 were withdrawn by the authors either before, during or after review.

Message from the Program Co-Chairs

point: we are growing rapidly, with corresponding heavy load on experienced researchers who participate in peer review; and we lack good tools to manage the process. Peer review involves several tasks that we, as NLP researchers, ought to be uniquely qualified to address, including expertise sourcing, network analysis and text mining. We have written a proposal with other members of the ACL community for concrete steps the ACL can take to improve our peer review infrastructure. We have also written a collection of “how to” documents that we will pass on to future conference organizers.

Ensure broad participation: To ensure broad participation, we recruited program committee members using a similar method to that used for NAACL HLT 2016: we invited anyone who had published repeatedly in ACL sponsored venues, who had a PhD or significant experience in the field spanning more than 5 years, and whose email address was up to date in START. We thank Dragomir Radev for giving us a list of names from the ACL anthology.

We also kept a blog where we discussed and attempted to “demystify” each stage of the publication process. This blog can be found at the conference website, <http://naacl2018.org>. We are very grateful to the researchers who wrote guest blog posts: Justine Cassell, Barbara Plank, Preslav Nakov, Omer Levy, Gemma Boleda, Emily Bender, Nitin Madnani, David Chiang, Kevin Knight, Dan Bikel and Joakim Nivre.

On our blog, we reported on the diversity of our area chair, reviewer and author pools in terms of years of experience, affiliation type and geography, and gender. We will include these details in our report to the NAACL Executive Committee. We hope that future years’ chairs will make similar reports.

The excellence of the overall NAACL HLT 2018 program is thanks to all the chairs and organizers. We especially thank the following people: Margaret Mitchell and Stephanie Lukin, the publication chairs; Jonathan May, the handbook chair; Yonatan Bisk and Wei Xu, the publicity and social media chairs; Ying Lin, the tireless website chair; and Marilyn Walker, the NAACL HLT 2018 general chair. We thank the chairs of NAACL HLT 2016 and ACL 2017 for their informative blogs, and the program chairs of NAACL HLT 2016, Owen Rambow and Ani Nenkova, for their advice. We thank the program co-chairs of ACL 2018, Iryna Gurevych and Yusuke Miyao, who have been very collaborative on matters related to reviewing. We thank Shuly Winter, who helped fix a serious START bug. We thank Julia Hockenmaier and the NAACL Executive Committee for their support. We are grateful for the professional work of Rich Gerber and his colleagues at SoftConf (START), and of Priscilla Rasmussen from the ACL.

It has been an enormous privilege for us to see the scientific advances that will be presented at this conference. We would like to close with some advice for you, the conference attendees.

- The presenters have made valuable contributions to our science; their oral, poster and demo presentations are worth your time and attention.
- Talk to some people you haven’t previously met. They may be your future collaborators!
- You can follow the conference on social media; we have a conference app and website where we will post any updates to the program, and our twitter handle is @naaclhlt.
- This event is run by a professional organization with a code of conduct (https://www.aclweb.org/adminwiki/index.php?title=Anti-Harassment_Policy). If you observe or are the recipient of unprofessional behavior, you may contact any current member of the ACL or NAACL Executive Committees, the NAACL HLT general chair (Marilyn Walker), us (the program chairs), or Priscilla Rasmussen (acl@aclweb.org). We will hold your communications in strict confidence and consult you before taking any action.

We look forward to a wonderful conference!

NAACL HLT 2018 Program Co-Chairs

Heng Ji, RPI

Amanda Stent, Bloomberg

Message from the Industry Track Co-Chairs

It is our pleasure to welcome you to the inaugural Industry Track in the *ACL family of conferences at NAACL HLT 2018.

The idea of organizing an industry track stemmed from the challenging issues encountered while attempting to apply state-of-the-art techniques to real-world language problems. As those who have attempted these problems know, practical applications are rarely as well defined as in laboratory settings and the data never as clean. In addition, there may be practical constraints such as computational requirements, processing speed, memory footprint, latency requirements, ease of updating a deployed solution that need to be balanced judiciously, and capability to be embedded as part of a larger system. The NAACL HLT 2018 Industry Track was born out of the desire to provide a forum for researchers, engineers and application developers who are experienced in these issues to exchange ideas, share results and discuss use-cases of successful development and deployments of language technologies in real-world settings.

Although we thought that the time is ripe in the NLP field for such a forum, and hoped that the community would embrace the opportunity to share their experience with others, it was nonetheless a guessing game as to the amount of interest the track would actually generate. As submissions drew to a close in late February, we were happy to report that we received 91 submissions, far exceeding our expectations (which led to last-minute scrambling to recruit more reviewers, but we're not complaining!). Six of the papers were desk rejects due to non-conformance with submission requirements, and the remaining 85 papers were reviewed by 65 reviewers. We accepted 28 papers – an acceptance rate of 32.9% (one paper was subsequently withdrawn after acceptance) of which 19 papers will be presented in oral sessions that run as a parallel track during the main conference, and 8 papers will be presented during poster sessions. Of course, none of this would have been possible without the participation of authors and reviewers, and we would like to convey our heartfelt “thank you” to all the authors for submitting papers and the reviewers for their efforts in the paper selection process.

We analyzed our submissions along a couple of dimensions and would like to share some interesting statistics. First we looked at the submissions with respect to the distribution of author affiliations. As one would expect, the industry track focuses on problems that manifest themselves more readily in industry than in academia. Indeed, of the 85 papers reviewed, 55 papers are authored by researchers/engineers in industry laboratories. The particularly encouraging statistic, however, is that 25 papers are the results of collaboration between those in industry and academia. It would be interesting to track these statistics in future years to see if the collaboration increases as the field continues to mature. The second dimension we analyzed is the geographic distribution of authors by contact author. This being a NAACL conference, it is no surprise that 62% of the papers came from North America. We are pleased with the participation of authors from other regions, including 22% from Europe, 14% from Asia, and 1% from Africa.

In addition to paper presentations, we will have two plenary keynote speeches. For the keynote speeches, we aimed to feature researchers who also have first hand experience applying research results to practical applications. To that end, we are honored to have two illustrious members of NLP community – Daniel Marcu, who co-founded Language Weaver more than 15 years ago and is now the director of MT/NLP at Amazon, and Mari Ostendorf, professor at the University of Washington, who led a team of students to build a social bot that won the 2017 Alexa Prize competition. We are confident that their experiences will be of immense interest to the larger NLP community.

Another highlight of the industry track includes two panel discussions on topics of increasing importance in the community. The first panel, “Careers in Industry,” moderated by Philip Resnik, professor at University of Maryland, is primarily geared toward students and recent graduates who are exploring careers in industry versus academia. The panel will feature experienced professionals who have worked in both environments to share their experience and offer advice, based on questions gathered from the *ACL community earlier this year. The second

Message from the Industry Track Co-Chairs

panel, “Ethics in NLP,” will be moderated by Dirk Hovy, professor at Università Bocconi, and will focus on raising awareness of the emerging issues of biases present in NLP/AI solutions, the social implications of such biases, and what we, as NLP practitioners, can do to reduce them.

With the overwhelming response to the call for papers, the language community has unanimously endorsed the relevance of the Industry track in the milieu of annual conferences. As organizers, we have attempted to amplify this endorsement by bringing to the participants an invigorating technical program. We hope through your engaging discussions and active participation during the sessions, you will unanimously support and nurture the concept of an Industry track in NLP conferences over the years to come.

Srinivas Bangalore (Interactions Labs)
Jennifer Chu-Carroll (Elemental Cognition)
Yunyao Li (IBM Research - Almaden)
NAACL HLT 2018 Industry Track Co-Chairs

Organizing Committee

General Chair

Marilyn Walker, University of California, Santa Cruz

Program Co-Chairs

Heng Ji, Rensselaer Polytechnic Institute

Amanda Stent, Bloomberg

Local Arrangements Chair

Priscilla Rasmussen, ACL

Workshop Co-Chairs

Marie Meteer, Brandeis University

Jason Williams, Microsoft Research

Industry Track Co-Chairs

Srinivas Bangalore, Interactions

Jennifer Chu-Carroll, Elemental Cognition

Yunyao Li, IBM

Tutorial Co-Chairs

Mohit Bansal, University of North Carolina

Rebecca Passonneau, Pennsylvania State University

Publication Co-Chairs

Stephanie Lukin, U.S. Army Research Laboratory

Margaret Mitchell, Google

Demonstration Co-Chairs

Yang Liu, University of Texas at Dallas

Tim Paek, Apple

Manasi Patwardhan, Tata Consultancy Services Research, India

Family-Friendly Program Co-Chairs

Chris Callison-Burch, University of Pennsylvania

Beth Hockey, Intel

Student Research Workshop Organizers

Silvio Ricardo Cordeiro, Aix-Marseille Université

Shereen Oraby, University of California, Santa Cruz

Umashanthi Pavalanathan, Georgia Institute of Technology

Kyeongmin Rim, Brandeis University

Faculty Advisors to the Student Research Workshop

Swapna Somasundaran, ETS Princeton

Sam Bowman, New York University

Publicity and Social Media Co-Chairs

Yonatan Bisk, University of Washington

Wei Xu, Ohio State University

Conference Handbook Chair

Jonathan May, University of Southern California / Information Sciences Institute

Student Volunteer Coordinator

Lena Reed, University of California, Santa Cruz

International Sponsorships Co-Chairs

Alexis Palmer, University of North Texas

Jason Baldridge, Google

Local Sponsorships and Exhibits Co-Chairs

Kristy Hollingshead, Florida Institute for Human & Machine Cognition

Kristen Johnson, Purdue University
Parisa Kordjamshidi, Tulane University, Florida Institute for Human & Machine Cognition

Website Chair

Ying Lin, Rensselaer Polytechnic Institute

Outstanding Paper Award Committee

Chair: Joel Tetreault, Grammarly

Joyce Chai, Michigan State University

Michael Collins, Columbia University

Jennifer Foster, Dublin City University

Smaranda Muresan, Columbia University

Test-of-Time Paper Expert Award Committee

Chair: Marilyn Walker, University of California, Santa Cruz

Antal van den Bosch, Radboud University Nijmegen

Chris Callison-Burch, University of Pennsylvania

Claire Cardie, Cornell University

Joyce Chai, Michigan State University

Walter Daelemans, University of Antwerp

Katrin Erk, University of Texas at Austin

Pascale Fung, The Hong Kong University of Science and Technology

Iryna Gurevych, TU Darmstadt

Katrin Kirchhoff, University of Washington

Yuji Matsumoto, Nara Institute of Science and Technology

Yusuke Miyao, National Institute of Informatics

Alessandro Moschitti, Qatar Computing Research Institute

Massimo Poesio, University of Essex

Owen Rambow, Elemental Cognition, LLC

Anoop Sarkar, Simon Fraser University

Amanda Stent, Bloomberg

Michael Strube, Heidelberg Institute for Theoretical Studies

Lucy Vanderwende, Microsoft Research

Hua Wu, Baidu Research

Chengqing Zong, Chinese Academy of Sciences

Program Committee

Program Co-Chairs

Heng Ji, Rensselaer Polytechnic Institute
Amanda Stent, Bloomberg

Area Chairs

Cognitive Modeling/Psycholinguistics

Morteza Dehghani, University of Southern California
Kristy Hollingshead Seitz, Institute for Human & Machine Cognition

Dialogue and Interactive Systems:

Yun-Nung (Vivian) Chen, National Taiwan University
Gabriel Skantze, KTH Royal Institute of Technology

Discourse and Pragmatics

Jacob Eisenstein, Georgia Institute of Technology
Junyi (Jessy) Li, University of Texas at Austin
Annie Louis, University of Edinburgh
Yi Yang, Bloomberg LP

Generation

Dimitra Gkatzia, Edinburgh Napier University
Shashi Narayan, University of Edinburgh
Michael White, Ohio State University

Information Extraction

Mausam, Indian Institute of Technology Delhi
Dan Bikel, Google
Chia-Hui Chang, National Central University
Bonan Min, BBN
Aurélie Névéol, CNRS
Marius Pasca, Google
Hinrich Schütze, Ludwig Maximilian University of Munich
Avirup Sil, IBM Research AI
Michael Strube, HITS gGmbH

Machine Learning for NLP

Chris Dyer, Google DeepMind
Ozan Irsoy, Bloomberg
Tie-Yan Liu, Microsoft Research
Raymond Mooney, University of Texas at Austin

Machine Translation

Marine Carpuat, University of Maryland
Kyunghyun Cho, New York University
Daniel Marcu, Amazon
Taro Watanabe, Google
Deyi Xiong, Soochow University

NLP Applications

Jinho Choi, Emory University
Joel Tetreault, Grammarly

Phonology, Morphology and Word Segmentation

Jennifer Foster, Dublin City University
Barbara Plank, University of Groningen

Question Answering

Eugene Agichtein, Emory University
Hannaneh Hajishirzi, University of Washington
Idan Szpektor, Google

Semantics

Yoav Artzi, Cornell University
Mona Diab, George Washington University
Kevin Duh, Johns Hopkins University
Jonathan May, University of Southern California / Information Sciences Institute
Preslav Nakov, Qatar Computing Research Institute
Roi Reichart, Technion - Israel Institute of Technology
Dan Roth, University of Pennsylvania
Scott Wen-tau Yih, Allen Institute for Artificial Intelligence

Sentiment Analysis

Smaranda Muresan, Columbia University
Swapna Somasundaran, ETS Princeton

Social Media Analysis and Computational Social Science

Mark Dredze, Johns Hopkins University
Miles Osborne, Bloomberg
Alan Ritter, Ohio State University
Sara Rosenthal, IBM
William Yang Wang, University of California, Santa Barbara

Speech

Eric Fosler-Lussier, Ohio State University
Dilek Hakkani-Tür, Google
Mari Ostendorf, University of Washington

Summarization

George Giannakopoulos, NCSR "Demokritos"
Xiaojun Wan, Peking University
Lu Wang, Northeastern University

Tagging, Chunking, Syntax and Parsing

Michael Collins, Columbia University
Yoav Goldberg, Bar Ilan University
Daisuke Kawahara, Kyoto University
Emily Pitler, Google
Anders Søgaard, University of Copenhagen
Aline Villavicencio, University of Essex (UK) and Federal University of Rio Grande do Sul (Brazil)

Text Mining

Kai-wei Chang, University of Virginia
Jing Jiang, Singapore Management University
Zornitsa Kozareva, Google
Chin-Yew Lin, Microsoft Research Asia

Theory and Formalisms

David Chiang, University of Notre Dame
Daniel Gildea, University of Rochester
Giorgio Satta, University of Padua

Vision, Robotics and Other Grounding

Joyce Chai, Michigan State University
Vicente Ordonez, University of Virginia

Outstanding Reviewers

We thank all reviewers for spending their valuable time on reviewing NAACL paper submissions! The following are outstanding reviewers nominated by area chairs:

- Omri Abend
- Manex Agirrezzabal
- Cem Akkaya
- Enrique Alfonseca
- Dimitrios Alikaniotis
- Tim Anderson
- Layla El Asri
- Michael Auli
- Wilker Aziz
- Mohit Bansal
- Valerio Basile
- Roberto Basili
- Beata Beigman Klebanov
- Jonathan Berant
- Delphine Bernhard
- Rahul Bhagat
- Bernd Bohnet
- Chris Brew
- Aoife Cahill
- Ruket Cakici
- Yuan Cao
- Giusepe Carenini
- Asli Celikyilmaz
- Cheng Chen
- Shamil Chollampatt
- Paul Crook
- Xiaodong Cui
- Giovanni da San Martino
- Andrew Dai
- Kees van Deemter
- Rodolfo Delmonte
- Gerard de Melo
- Dina Demner-Fushman
- Lingjia Deng
- Mark Dras
- Ondřej Dušek
- Marc Dymetman
- Steffen Eger
- Markus Egg
- Keelan Evanini
- Olivier Ferret
- Katja Filippova
- Michael Flor
- Anette Frank
- Michel Galley
- Claire Gardent
- Kallirroi Georgila
- Yvette Graham
- Debanjan Ghosh
- Francisco Guzman
- Thierry Hamon
- Helen Hastie
- Michael Heilman
- James Hender
- Derrick Higgins
- Chiori Hori
- Veronique Hoste
- Julian Hough
- Hen-Hsen Huang
- Rebecca Hwa
- Marcin Junczys-Dowmunt
- Rasoul Kaljahi
- Lauri Karttunen
- David Kauchak
- Anna Kazantzева
- Andrew Kehler
- Kian Kenyon-Dean
- Alistair Knott
- Hayato Kobayashi
- Mamoru Komachi
- Alexander Koller
- Patrik Lambert
- Mirella Lapata
- John Lee
- Yves Lepage
- James Lester
- Omer Levy
- Yanran Li
- Chu-Cheng Lin
- Elena Lloret
- Karen Livescu
- Bing Liu
- Stephanie Lukin
- Nitin Madnani
- Adam Makarucha
- Yoichi Matsuyama
- Karen Mazidi
- Diana McCarthy
- Philipp Meerkamp
- Oren Melamud
- Emiel Van Miltenburg
- Saif Mohammad
- Taesun Moon
- Nasrin Mostafazadeh
- Philippe Muller
- Karthik Narasimhan
- Dong Nguyen
- Garrett Nicolai
- Hugo Gonçalo Oliveira
- Lilja Øvreliid
- Alexis Palmer
- Laura Perez-Beltrachini
- Paul Piwek
- Thierry Poibeau
- Fred Popowich
- Daniel Preotiu-Pietro
- Carlos Ramisch
- Sujith Ravi
- Paul Rayson
- Sravana Reddy
- Marek Rei
- Ehud Reiter
- Ludovic Rheault
- Giuseppe Riccardi
- Mark Riedl
- Laura Rimel
- Brian Roark
- Marcus Rohrbach
- Melissa Roemmel
- Eric Rosen
- Johann Roturier
- Markus Saers
- Allen Schmaltz
- Jodi Schneider
- H. Andrew Schwartz
- Roy Schwartz
- Donia Scott
- Advaith Siddharthan
- Michel Simard
- Kairit Sirts
- Jose de Souza
- Somayajulu Sripada
- Matthew Stone
- Kristina Striegnitz
- John Sylak-Glassman
- Kapil Thadani
- Mariët Theune
- Timothy Baldwin
- Antonio Toral
- Marco Turchi
- Andreas Vlachos
- Leo Wanner
- Michael Wiegand
- Shuly Winter
- Sam Wiseman
- Jiajun Zhang
- Arkaitz Zubiaga

Ad-hoc Reviewers

The following are our heroes and heroines! They volunteered their time when we were painfully chasing late reviews.

- Cem Akkaya
- Héctor Martínez Alonso
- Waleed Ammar
- Gabor Angeli
- Reihaneh Boghrati
- Iacer Calixto
- Leonardo Campillos
- Kai Cao
- Özlem Çetinoğlu
- Yee Seng Chan
- Michael Collins
- Çağrı Çöltekin
- Silvio Ricardo Cordeiro
- Bo Dai
- Noura Farra
- Michael Flor
- Radu Florian
- Justin Garten
- Tao Ge
- Debanjan Ghosh
- Jiatao Gu
- Keith Hall
- Benjamin Heinzerling
- Xianpei Han
- Rasoul Kaljahi
- Philipp Koehn
- Jonathan K. Kummerfeld
- Xiang Li
- Elena Lloret
- Bing Liu
- Jose David Lopes
- Di Lu
- Nikolaos Malandrakis
- Jeff Mitchell
- Saif Mohammad
- Matthew Mulholand
- Vivi Nastase
- Dong Nguyen
- Thienn Huu Nguyen
- Hugo Gonçalo Oliveira
- Shereen Oraby
- Vinodkumar Prabhakaran
- Sujith Ravi
- Sophie Rosset
- Natalie Schluter
- Roy Schwartz
- Lanbo She
- Amanda Stent
- Veselin Stoyanov
- Keh-Yih Su
- Chen-Tse Tsai
- Gokhan Tür
- Svitlana Volkova
- Joachim Wagner
- Chi Wang
- Taro Watanabe
- Chris Waterson
- Rodrigo Wilkens
- Mark Yatskar
- Wenpeng Yin
- Boliang Zhang
- Arkaitz Zubiaga

Meal Info

The following meals are provided as part of your registration fee:

- A full buffet breakfast will be provided each day from 07:30–09:00 in the Empire Foyer
- Mid-morning breaks include coffee and tea in the Empire Foyer
- Mid-afternoon breaks include coffee, tea, soda, water, and snacks in the Empire Foyer
- The Welcome Reception on Friday, June 1st, from 18:00–21:00, will include a light dinner
- The Recruitment Lunch on Saturday, June 2nd, 12:30–14:00, is intended for students and recruiters and will include a light lunch
- A full dinner buffet is provided during the Social Event on Sunday, June 3rd

2

Workshops and Tutorials: Friday, June 1

Overview

07:30 – 18:00	Registration	<i>Empire Foyer</i>
07:30 – 9:00	Breakfast	<i>Empire Foyer</i>
08:30 – 18:00	WiNLP: Widening NLP	<i>Strand 12</i>
09:00 – 12:30	Morning Tutorials	
	Modelling Natural Language, Programs, and their Intersection <i>Graham Neubig and Miltiadis Allamanis</i>	<i>Bolden 5</i>
	Deep Learning Approaches to Text Production <i>Claire Gardent and Shashi Narayan</i>	<i>Bolden 6</i>
	Scalable Construction and Reasoning of Massive Knowledge Bases <i>Xiang Ren, Nanyun Peng, and William Yang Wang</i>	<i>Strand 13</i>
10:00 – 10:30	Coffee break	
12:30 – 14:00	Lunch break	
14:00 – 17:30	Afternoon Tutorials	
	The Interplay Between Lexical Resources and Natural Language Processing <i>Jose Camacho-Collados, Luis Espinosa Anke, and Mohammad Taher Pilehvar</i>	<i>Bolden 5</i>
	Socially Responsible NLP <i>Yulia Tsvetkov, Vinodkumar Prabhakaran, and Rob Voigt</i>	<i>Bolden 6</i>
	Deep Learning for Conversational AI <i>Pei-Hao Su, Nikola Mrkšić, Iñigo Casanueva, and Ivan Vulić</i>	<i>Strand 13</i>
15:00 – 15:30	Coffee break	
18:00 – 21:00	Welcome Reception	<i>Empire Ballroom</i>

WiNLP: Widening NLP

Organizers: *Libby Barak, Amitai Axelrod, Lucie Flekova, Diyi Yang, and Zeerak Waseem*

Venue: Strand 12

Friday, June 1, 2018

Session I: Opening, Invited Talk I, and Oral Presentations

07:30–18:00 **NAACL Registration**

08:30–08:50 **Opening Remarks:** *Libby Barak, Amitai Axelrod, Diyi Yang, Lucie Flekova, Zeerak Waseem*

08:55–09:30 **Invited Talk I: Su Jian - Semantic and Sentiment Analysis for Knowledge Base Population**

09:30–10:30 **Oral Presentations**

09:30–09:50 Incorporating Subword Information into Matrix Factorization Word Embeddings
Alexandre Salle and Aline Villavicencio

09:50–10:10 #MeToo Alexa: Sexual Harassment and Conversational AI
Amanda Cercas Curry and Verena Rieser

10:10–10:30 Multi-label Element Extraction for Evidence-Based Medicine
Amandalynne Paullada

10:30–11:00 **Coffee Break**

Session II: Invited Talk, Mentoring

11:00–11:35 **Invited Talk II: Natalie Schluter - The glass ceiling in NLP**

11:35–12:00 **Mentoring**

12:00–13:00 **Lunch**

Session III: Poster Presentations, Career Panel

13:00–14:30 **Poster Presentations**

13:00–14:30 Luganda Text-to-Speech Machine
Irene Nandutu and Ernest Mwebaze

13:00–14:30 Context-Sensitive Normalization of Social Media Text in Bahasa Indonesia Based on Neural Word Embeddings
Renny Pradina Kusumawardani and Stezar Priansya

13:00–14:30 A novel Post-editing algorithm to optimize Amharic speech recognition for speech translation
Michael Melese, Laurent Besacier, and Million Meshesha

13:00–14:30 Building on Word Animacy to Determine Coreference Chain Animacy in Cultural Narratives
Labiba Jahan, Geeticka Chauhan, and Mark A. Finlayson

13:00–14:30 Neural Morphological Segmentation for Polysynthetic Minimal-Resource Languages
Katharina Kann, Manuel Mager, Ivan Vladimir Meza Ruiz, and Hinrich Schütze

- 13:00–14:30 Empirical Evaluation of Preprocessing for Tunisian Dialect ("Darija")
Sentiment Analysis
Hatem Haddad and Hala Mulki
- 13:00–14:30 Human vs Automatic Metrics: on the Importance of Correlation Design
Anastasia Shimorina
- 13:00–14:30 Sentence Level Amharic Text Sentiment Analysis Model: A Combined Approach
Bitseat Aragaw, Tensae Ayalew, and Tibebe Beshah
- 13:00–14:30 Development of Pronunciation Lexicons for Amharic Automatic Speech Recognition (ASR)
Martha Yifru Tachbelie and Solomon Teferra Abate
- 13:00–14:30 Neural Network-based Model for Named Entity Recognition in Low-Resource Settings
Ngoc Tan Le, Hong Buu Long Nguyen, Fatiha Sadat, and Dien Dinh
- 13:00–14:30 Data Adaptation for Named Entity Recognition in Twitter with Features-Rich CRF
Ngoc Tan Le and Fatiha Sadat
- 13:00–14:30 Linguistic Accommodation and Codeswitching in Conversation
Rouzbeh Shirvani, Mario Piergallini, Amirsina Torfi, Mugizi Rwebangira, and Mohamed Chouikha
- 13:00–14:30 Improving Zero-Shot Translation of Low-Resource Languages
Surafel Melaku Lakew, Quintino Francesco Lotito, Matteo Negri, Marco Turchi, and Marcello Federico
- 13:00–14:30 Emotionality in Patients and Therapists speaking German
Margot Mieskes and Svetlana Shutyi
- 13:00–14:30 Concept Linking in Medical Domain for Social Media Posts Using RNNs
Elena Tutubalina, Zulfat Miftahutdinov, and Valentin Malykh
- 13:00–14:30 Predicting Glosses in Interlinear Glossed Text
Angelina McMillan-Major
- 13:00–14:30 The Entrainment of Creaky Voice in Spoken Dialogue
Courtney Mansfield
- 13:00–14:30 NeuroSign: a Seq2Seq Model Proposal for Sign Language Translation
Héctor Ricardo Murrieta Bello and Ivan Vladimir Meza Ruiz
- 13:00–14:30 Multimodal Multilingual Representation Learning
Spandana Gella
- 13:00–14:30 #MeToo: Neural Detection and Explanation of Language in Personal Abuse Stories
Sweta Karlekar and Mohit Bansal
- 13:00–14:30 Text Mining of Student Generated Content for Culturally Relevant Pedagogies in Data Science
Jaye Nias
- 13:00–14:30 A Corpus for the Analysis of Online News Comments
Varada Kolhatkar and Maite Taboada
- 13:00–14:30 Modeling of Policy Frames for Morality Detection on Twitter
Kristen Johnson and Dan Goldwasser
- 13:00–14:30 Longer Acoustic Units for Amharic Speech Recognition
Adey Edessa and Martha Yifru
- 13:00–14:30 A Framework for Suggestion Mining from Vietnamese Conversational Texts
Thi Lan Ngo and Pham Khac Linh
- 13:00–14:30 Fluency Over Adequacy: A Pilot Study in Measuring User Trust in Imperfect MT
Marianna Martindale and Marine Carpuat
- 13:00–14:30 Using Machine Learning to Detect Potential Child Suicide Bombers
Francisca Oladipo and Haruna Abdu

- 13:00–14:30 Learning Diacritic Embeddings
Ignatius Ezeani, Ikechukwu Onyenwe, Mark Hepple, and Enemouh Chioma
- 13:00–14:30 Transfer between Disparate Labels with Weak Supervision
Rasha Obeidat, Xiaoli Fern, and Prasad Tadepalli
- 13:00–14:30 Unsupervised Sentence Embeddings for Answer Summarization in Non-factoid CQA
Thi-Thanh Ha, Thanh-Chinh Nguyen, Kiem-Hieu Nguyen, and Van-Chung Vu
- 13:00–14:30 New Entity Identification Approaches in Entity Linking Systems
Priya Radhakrishnan, Manish Gupta, and Vasudeva Varma
- 13:00–14:30 A finite-state morphological analyzer for Wolayta
Tewodros Gebreselassie
- 13:00–14:30 Exploring Morphology Learning of Igbo with Different Degrees of Supervision
Enemouh Chioma, Ignatius Ezeani, Mark Hepple, and Ikechukwu Onyenwe
- 13:00–14:30 Comparing the Performance of Crowdworkers and NLP Tools on Named-Entity Recognition and Entity-Level Sentiment Analysis of Political Tweets
Mona Jalal, Kate K. Mays, Lei Guo, and Margrit Betke
- 13:00–14:30 Implicit and Explicit Information for Categorization
Felipe Paula, Rodrigo Wilkens, Marco Idiart, and Aline Villavicencio
- 13:00–14:30 Computational Analysis of Polarity in Poetry
Navjyoti Singh and Amitha Gopidi
- 13:00–14:30 Differences in the Acquisition of Polysemy vs. Ambiguity
Sammy Floyd, Libby Barak, and Adele Goldberg
- 13:00–14:30 Language Models for Brain Computer Interface as Augmentative and Alternative Communication
Shiran Dudy, Steven Bedrick, David Smith, and Shaobin Xu
- 13:00–14:30 Lightweight Word-Level Confidence Estimation for Neural Interactive Translation Prediction
Rebecca Knowles and Philipp Koehn
- 13:00–14:30 Hierarchical Neural Model for Tagging Address Queries in Map Search
Shekoofeh Mokhtari, Ning Xie, and Dragomir Yankov
- 13:00–14:30 Sentiment Based Word Substitutions to Generate Humorous Phrases in Hindi
Srishti Aggarwal, Kritik Mathur, and Radhika Mamidi
- 13:00–14:30 Faster Pace in Human-Robot Dialogue Leads to Fewer Dialogue Overlaps
Cassidy Henry, Carla Gordon, David Traum, Stephanie Lukin, Kimberly A. Pollard, Ron Artstein, Claire Bonial, Clare Voss, Ashley Foots, and Matthew Marge
- 13:00–14:30 Enrichment of OntoSenseNet: Adding a Sense-annotated Telugu Lexicon
Sreekavitha Parupalli and Navjyoti Singh
- 13:00–14:30 Cross-Situational Word Learning under Uncertainty: an Algorithmic-Level Analysis
Aida Nematzadeh, Zahra Shekarchi, and Suzanne Stevenson
- 13:00–14:30 Generating Creative Restaurant Descriptions from User Reviews
Shereen Oraby and Marilyn Walker
- 13:00–14:30 Learning Lexico-Functional Patterns for First-Person Affect
Lena Reed and Marilyn Walker
- 13:00–14:30 Building a NLI Corpus for Portuguese
Lívia Real, Valéria de Paiva, Bruno Guide, Bruna Thalenberg, Cindy Silva, Guilherme Oliveira Lima, Ana Rodrigues, Beatriz Albiero, Rodrigo Souza, and Igor de Camargo e Souza Câmara
-

13:00–14:30 **WitWise: A Funny Proverb Generator**
Geetanjali Rakshit

14:30–15:30 **Career Panel: Houda Bouamor, Barbara Plank, Mugizi Robert Rwebangira, and Amanda Stent**

15:30–15:55 **Coffee Break**

Session IV: Invited Talk, Oral Presentations, Closing, Mentor Meetup

15:55–16:30 **Invited Talk III: Claire Cardie - Information Extraction: What Goes Around Comes Around**

16:30–17:10 **Oral Presentations**

16:30–16:50 Churn Prediction Using Structured Logical Knowledge and Convolutional Neural Networks

Mourad Gridach and Hatem Haddad

16:50–17:10 Creating and Annotating a Corpus of Health Coaching Dialogue
Itika Gupta, Barbara Di Eugenio, Brian Ziebart, Bing Liu, Ben Gerber, Lisa Sharp, Rafe Davis, and Aiswarya Baiju

17:10–17:30 **Closing Remarks - Libby Barak, Amitai Axelrod, Diyi Yang, Lucie Flekova, Zeerak Waseem**

17:30–18:00 **One-on-One Mentor Meetup**

18:00 **NAACL Welcome Reception**

Message from the Tutorial Co-Chairs

Welcome to the Tutorials Session of NAACL HLT 2018 in New Orleans.

The NAACL HLT tutorials session is an opportunity for conference attendees to participate in a tutorial on a timely topic of importance to the field, and to hear from experts on that topic. This year, the tutorials committee comprised tutorials chairs from four conferences: ACL, COLING, EMNLP and NAACL HLT. A total of 51 tutorial submissions were received, of which 6 were selected for presentation at NAACL HLT.

We hope you find the six NAACL HLT tutorials for this year to combine depth within each tutorial, and breadth across a set of topics that demonstrate the increasing relevance of NLP to a wide range of fields within and beyond computer science.

We would like to thank Marilyn Walker (NAACL general chair), Ying Lin (NAACL website chair), Stephanie Lukin and Margaret Mitchell (NAACL publications chairs), and Priscilla Rasmussen (local arrangement chair) for their help during the whole process. We also want to extend our sincere gratitude to the other conferences' tutorial chairs who jointly helped with reviewing for all the tutorial submissions: Yoav Artzi, Jacob Eisenstein, Pascale Fung, Donia Scott, Marilyn Walker, Mausam, and Lu Wang.

Enjoy the tutorials!

NAACL 2018 Tutorial Co-chairs

Mohit Bansal

Rebecca Passonneau

Tutorial 1

Modelling Natural Language, Programs, and their Intersection

Graham Neubig and Miltiadis Allamanis

Friday, June 1, 2018, 09:00–12:30

Bolden 5

As computers and information grow a more integral part of our world, it is becoming more and more important for humans to be able to interact with their computers in complex ways. One way to do so is by programming, but the ability to understand and generate programming languages is a highly specialized skill. As a result, in the past several years there has been an increasing research interest in methods that focus on the intersection of programming and natural language, allowing users to use natural language to interact with computers in the complex ways that programs allow us to do. In this tutorial, we will focus on machine learning models of programs and natural language focused on making this goal a reality. First, we will discuss the similarities and differences between programming and natural language. Then we will discuss methods that have been designed to cover a variety of tasks in this field, including automatic explanation of programs in natural language (code-to-language), automatic generation of programs from natural language specifications (language-to-code), modeling the natural language elements of source code, and analysis of communication in collaborative programming communities. The tutorial will be aimed at NLP researchers and practitioners, aiming to describe the interesting opportunities that models at the intersection of natural and programming languages provide, and also how their techniques could provide benefit to the practice of software engineering as a whole.

Graham Neubig (gneubig@cs.cmu.edu, <http://phontron.com>) is an assistant professor at Carnegie Mellon University specializing in natural language processing and machine learning. One of his major research interests is models that link together natural language and code, including summarizing the intent of code in natural language, generating code from natural language, or discovering the correspondences between the two modalities. He has previously given well-attended tutorials at NLP conferences (EMNLP and YRSNLP) and the Lisbon Machine Learning Summer School, and has won a number of best papers (e.g. EMNLP2016 and EACL2017) and given invited talks, including an upcoming one on this topic at the AAAI Workshop on NLP for Software Engineering.

Miltiadis Allamanis (miallama@microsoft.com, <https://miltos.allamanis.com>) is a researcher at Microsoft Research, Cambridge, UK at the Deep Program Understanding project. He is researching applications of machine learning and natural language processing to software engineering and programming languages to create smart software engineering tools for developers. Miltos has published in both machine learning and software engineering conferences and is an author of a recent survey on machine learning for source code (<https://ml4code.github.io>). He received his PhD at the University of Edinburgh, UK advised by Dr. Charles Sutton.

Tutorial 2

Deep Learning Approaches to Text Production

Claire Gardent and Shashi Narayan

Friday, June 1, 2018, 09:00–12:30

Bolden 6

Text production is a key component of many NLP applications. In data-driven approaches, it is used for instance, to generate dialogue turns from dialogue moves, to verbalise the content of Knowledge bases or to generate natural English sentences from rich linguistic representations, such as dependency trees or Abstract Meaning Representations. In text-driven methods on the other hand, text production is at work in sentence compression, sentence fusion, paraphrasing, sentence (or text) simplification, text summarisation and end-to-end dialogue systems.

Following the success of encoder-decoder models in modeling sequence-rewriting tasks such as machine translation, deep learning models have successfully been applied to the various text production tasks. In this tutorial, we will cover the fundamentals and the state-of-the-art research on neural models for text production. Each text production task raises a slightly different communication goal (e.g, how to take the dialogue context into account when producing a dialogue turn; how to detect and merge relevant information when summarising a text; or how to produce a well-formed text that correctly capture the information contained in some input data in the case of data-to-text generation). We will outline the constraints specific to each subtasks and examine how the existing neural models account for them.

Claire Gardent (claire.gardent@loria.fr, <https://members.loria.fr/CGardent/>) is a research scientist at CNRS (the French National Center for Scientific Research). Prior to joining the CNRS, she worked at the Université de Clermont-Ferrand, Saarbrücken Universität and Amsterdam Universiteit. She received her Ph.D. degree from the University of Edinburgh. Her research interests include (executable) semantic parsing, natural language generation and simplification and, more recently, the use of computational linguistics for linguistic analysis. She was nominated Chair of the EACL and acted as program chair for various international conferences, workshops and summer schools (EACL, ENLG, SemDIAL, SIGDIAL, ESSLLI, *SEM). She currently heads the WebNLG project (Nancy, Bolzano, Stanford SRI) and is the chair of SIGGEN, the ACL Special Interest Group in Natural Language Generation. Recently she co-organised the WebNLG Shared Task, a challenge on generating text from RDF data.

Shashi Narayan (shashi.narayan@ed.ac.uk, <http://homepages.inf.ed.ac.uk/snaraya2/>) is a research associate at the School of Informatics at the University of Edinburgh. His research focuses on natural language generation, understanding and structured predictions. A major aim of his research is to build on the hypothesis that tailoring a model with knowledge of the task structure and linguistic requirements, such as syntax and semantics, leads to a better performance. The questions raised in his research are relevant to various natural language applications such as question answering, paraphrase generation, semantic and syntactic parsing, document understanding and summarization, and text simplification. He mostly rely on machine learning techniques such as deep learning and spectral methods to develop NLP frameworks. His research has appeared in computational linguistics journals (e.g., TACL, Computational Linguistics and Pattern Recognition Letters) and in conferences (e.g., ACL, EMNLP, COLING, EACL and INLG). He was nominated on the SIGGEN board (2012-14) as a student member. He co-organised the WebNLG Shared Task, a challenge on generating text from RDF data. Recently, he is nominated as an area co-chair for Generation at NAACL HLT 2018.

Tutorial 3

Scalable Construction and Reasoning of Massive Knowledge Bases

Xiang Ren, Nanyun Peng, and William Yang Wang

Friday, June 1, 2018, 09:00–12:30

Strand 13

In today’s information-based society, there is abundant knowledge out there carried in the form of natural language texts (e.g., news articles, social media posts, scientific publications), which spans across various domains (e.g., corporate documents, advertisements, legal acts, medical reports), which grows at an astonishing rate. Yet this knowledge is mostly inaccessible to computers and overwhelming for human experts to absorb. How to turn such massive and unstructured text data into structured, actionable knowledge, and furthermore, how to teach machines learn to reason and complete the extracted knowledge is a grand challenge to the research community.

Traditional IE systems assume abundant human annotations for training high quality machine learning models, which is impractical when trying to deploy IE systems to a broad range of domains, settings and languages.

In the first part of the tutorial, we introduce how to extract structured facts (i.e., entities and their relations for types of interest) from text corpora to construct knowledge bases, with a focus on methods that are weakly-supervised and domain-independent for timely knowledge base construction across various application domains.

In the second part, we introduce how to leverage other knowledge, such as the distributional statistics of characters and words, the annotations for other tasks and other domains, and the linguistics and problem structures, to combat the problem of inadequate supervision, and conduct low-resource information extraction.

In the third part, we describe recent advances in knowledge base reasoning. We start with the gentle introduction to the literature, focusing on path-based and embedding based methods. We then describe DeepPath, a recent attempt of using deep reinforcement learning to combine the best of both worlds for knowledge base reasoning.

Xiang Ren (xiangren@usc.edu, <http://www.bcf.usc.edu/~xiangren/>). Assistant Professor, Department of Computer Science, University of Southern California. His research focuses on creating computational tools for better understanding and exploring massive text data. He has published over 25 papers in major conferences. He received Google PhD Fellowship, KDD Rising Star by Microsoft, Yahoo!-DAIS Research Excellence Award, C. W. Gear Outstanding Graduate Student Award by UIUC and Yelp Dataset Challenge Award. Mr. Ren has rich experiences in delivering tutorials in major conferences, including SIGKDD 2015, SIGMOD 2016 and WWW 2017.

Nanyun Peng (npeng@isi.edu, <http://www.vnpeng.net>), Computer Scientist, Information Sciences Institute, University of Southern California. She is broadly interested in Natural Language Processing, Machine Learning, and Information Extraction. Her research focuses on low-resource information extraction, creative language generation, and phonology/morphology modeling. Nanyun is the recipient of the Johns Hopkins University 2016 Fred Jelinek Fellowship. She has a background in computational linguistics and economics and holds BAs in both.

William Yang Wang (wiliam@cs.ucsb.edu, <http://www.cs.ucsb.edu/~william/>) is an Assistant Professor at the Department of Computer Science, University of California, Santa Barbara. He received his PhD from Carnegie Mellon University, where he worked on scalable probabilistic reasoning language ProPPR with William Cohen. He focuses on information extraction and he is the faculty author of DeepPath—the first deep reinforcement learning system for multi-hop knowledge reasoning. He has published more than 40 papers at leading conferences and journals including ACL, EMNLP, NAACL, COLING, IJCAI, CIKM, SIGDIAL, IJCNLP, INTERSPEECH, ICASSP, ASRU, SLT, Machine Learning, and Computer Speech & Language, and he has received paper awards and honors from CIKM, ASRU, and EMNLP.

Tutorial 4

The Interplay Between Lexical Resources and Natural Language Processing

Jose Camacho-Collados, Luis Espinosa Anke, and Mohammad Taher Pilehvar

Friday, June 1, 2018, 14:00–17:30

Bolden 5

Incorporating linguistic, world and common sense knowledge into AI/NLP systems is currently an important research area, with several open problems and challenges. At the same time, processing and storing this knowledge in lexical resources is not a straightforward task. We propose to address these complementary goals from two methodological perspectives: the use of NLP methods to help the process of constructing and enriching lexical resources and the use of lexical resources for improving NLP applications. This tutorial may be useful for two main types of audience: those working on language resources who are interested in becoming acquainted with automatic NLP techniques, with the end goal of speeding and/or easing up the process of resource curation; and on the other hand, researchers in NLP who would like to benefit from the knowledge of lexical resources to improve their systems and models.

Jose Camacho-Collados (collados@di.uniroma1.it, <http://www.josecamachocollados.com>) is a Google Doctoral Fellow and PhD student at Sapienza University of Rome. His research focuses on Natural Language Processing and, more specifically, on the area of lexical and distributional semantics. Jose has experience in utilizing lexical resources for NLP applications, while enriching and improving these resources by extracting and processing knowledge from textual data. On this area he is co-organizing the upcoming SemEval 2018 shared task on Hypernym Discovery. Previously, he co-organized a workshop on “Sense, Concept and Entity Representations and their Applications” at EACL 2017 and a tutorial on the same topic at ACL 2016. His background education includes an Erasmus Mundus Master in Natural Language Processing and Human Language Technology and a 5-year BSc degree in Mathematics.

Luis Espinosa Anke (espinosa-anke@cardiff.ac.uk, <http://www.luisespinoza.net>) received his BA in English Philology in 2006 (Univ. of Alicante, Spain), and his PhD in Natural Language Processing in 2017 (Univ. Pompeu Fabra, Spain). He holds two MAs, one in English-Spanish Translation (Univ. of Alicante), and an Erasmus Mundus MA in Natural Language Processing (NLP) (Univ. of Wolverhampton and Univ. Autònoma de Barcelona). His research interests lie in the intersection between structured representations of knowledge and NLP, specifically computational lexicography and distributional semantics. He is a co-organizer of the upcoming SemEval 2018 shared tasks on Hypernym Discovery and Multilingual Emoji Prediction. Previously, he co-organized the Spanish NLP conference (2014), the Focused NER task at the Open Knowledge Extraction challenge at ESWC (2017).

Mohammad Taher Pilehvar (mp792@cam.ac.uk, <http://people.ds.cam.ac.uk/mp792/>) is a research associate at the University of Cambridge. Taher’s research lies in lexical semantics, mainly focusing on semantic representation and similarity. In the past, he has co-instructed three tutorials on these topics (EMNLP 2015, ACL 2016, and EACL 2017) and co-organised three SemEval tasks. He has also co-authored several conference (including two ACL best paper nominations, at 2013 and 2017) and journal papers, including different semantic representation techniques based on heterogeneous lexical resources.

Tutorial 5

Socially Responsible NLP

Yulia Tsvetkov, Vinodkumar Prabhakaran, and Rob Voigt

Friday, June 1, 2018, 14:00–17:30

Bolden 6

As language technologies have become increasingly prevalent, there is a growing awareness that decisions we make about our data, methods, and tools are often tied up with their impact on people and societies. This tutorial will provide an overview of real-world applications of language technologies and the potential ethical implications associated with them. We will discuss philosophical foundations of ethical research along with state of the art techniques. Through this tutorial, we intend to provide the NLP researcher with an overview of tools to ensure that the data, algorithms, and models that they build are socially responsible. These tools will include a checklist of common pitfalls that one should avoid (e.g., demographic bias in data collection), as well as methods to adequately mitigate these issues (e.g., adjusting sampling rates or de-biasing through regularization). The tutorial is based on a new course on Ethics and NLP developed at Carnegie Mellon University.

Yulia Tsvetkov (ytsvetko@cs.cmu.edu, <http://www.cs.cmu.edu/~ytsvetko/>) is an assistant professor in the Language Technologies Institute at Carnegie Mellon University. Her research interests lie at or near the intersection of natural language processing, machine learning, linguistics, and social science. Her current research projects focus on NLP for social good, including advancing language technologies for resource-poor languages spoken by millions of people, developing approaches to promote civility in communication (e.g., modeling gender bias in texts and debiasing), identifying strategies that undermine the democratic process (e.g., political framing and agenda-setting in digital media). Prior to joining CMU, Yulia was a postdoc in the Stanford NLP Group; she received her PhD from Carnegie Mellon University.

Vinodkumar Prabhakaran (vinod@cs.stanford.edu, www.cs.stanford.edu/~vinod) is a postdoctoral fellow at the Stanford NLP lab, and prior to this, received his PhD in Computer Science from Columbia University in 2015. His research falls in the interdisciplinary field of computational social sciences, with a focus on applying NLP for social good. He combines NLP techniques with social science methods in order to identify and address large scale societal issues, such as racial bias and disparities in law enforcement, manifestations of power and gender at workplace, and online incivility such as condescension and gender bias.

Rob Voigt (robvoigt@stanford.edu, <https://nlp.stanford.edu/robvoigt/>) is a PhD student in the Linguistics Department at Stanford University, working on topics in computational sociolinguistics with Dan Jurafsky. His research focuses on using computational methods to understand how social context and social factors subtly influence linguistic behavior at a large scale. His dissertation is focused on techniques for extracting and analyzing linguistic implicit bias, including respectfulness in policecommunity interaction, gender bias in online communications, and “othering” in historical media representations of immigrant groups.

Tutorial 6

Deep Learning for Conversational AI

Pei-Hao Su, Nikola Mrkšić, Iñigo Casanueva, and Ivan Vulić

Friday, June 1, 2018, 14:00–17:30

Strand 13

Spoken Dialogue Systems (SDS) have great commercial potential as they promise to revolutionise the way in which humans interact with machines. The advent of deep learning led to substantial developments in this area of NLP research, and the goal of this tutorial is to familiarise the research community with the recent advances in what some call the most difficult problem in NLP.

From a research perspective, the design of spoken dialogue systems provides a number of significant challenges, as these systems depend on: a) solving several difficult NLP and decision-making tasks; and b) combining these into a functional dialogue system pipeline. A key long-term goal of dialogue system research is to enable open-domain systems that can converse about arbitrary topics and assist humans with completing a wide range of tasks. Furthermore, such systems need to autonomously learn on-line to improve their performance and recover from errors using both signals from their environment and from implicit and explicit user feedback. While the design of such systems has traditionally been modular, domain and language-specific, advances in deep learning have alleviated many of the design problems.

The main purpose of this tutorial is to encourage dialogue research in the NLP community by providing the research background, a survey of available resources, and giving key insights to application of state-of-the-art SDS methodology into industry-scale conversational AI systems. We plan to introduce researchers to the pipeline framework for modelling goal-oriented dialogue systems, which includes three key components: 1) Language Understanding; 2) Dialogue Management; and 3) Language Generation. The differences between goal-oriented dialogue systems and chat-bot style conversational agents will be explained in order to show the motivation behind the design of both, with the main focus on the pipeline SDS framework. For each key component, we will define the research problem, provide a brief literature review and introduce the current state-of-the-art approaches. Complementary resources (e.g. available datasets and toolkits) will also be discussed. Finally, future work, outstanding challenges, and current industry practices will be presented. All of the presented material will be made available online for future reference.

Pei-Hao (Eddie) Su (eddysu@poly-ai.com, <https://eddy0613.github.io/>) is a co-founder and CTO of PolyAI, a London-based startup looking to use the latest developments in NLP to create a general machine learning platform for deploying spoken dialogue systems. He holds a PhD from the Dialogue Systems group, University of Cambridge, where he worked under the supervision of Professor Steve Young. His research interests centre on applying deep learning, reinforcement learning and Bayesian approaches to dialogue management and reward estimation, with the aim of building systems that can learn directly from human interaction. He has given several invited talks at academia and industry such as Apple, Microsoft, General Motor and DeepHack.Turing. He received the best student paper award at ACL 2016.

Nikola Mrkšić (nikola@poly-ai.com, <http://mi.eng.cam.ac.uk/nm480>) a co-founder and CEO of PolyAI, a London-based startup looking to use the latest developments in NLP to create a general machine learning platform for deploying spoken dialogue systems. He holds a PhD

from the Dialogue Systems group, University of Cambridge, where he worked under the supervision of Professor Steve Young. His research is focused on belief tracking in human-machine dialogue, specifically in moving towards building open-domain, cross-lingual language understanding models that are fully data-driven. He is also interested in deep learning, semantics, Bayesian nonparametrics, unsupervised and semi-supervised learning. He previously gave a tutorial on word vector space specialisation at EACL 2017, and will teach a course on the same topic at ESSLLI 2018. He also gave invited talks at the REWORK AI Personal Assistant summit and the Chatbot Summit.

Inigo Casanueva (inigo@poly-ai.com, [http://mi.eng.cam.ac.uk/ ic340/](http://mi.eng.cam.ac.uk/ic340/)) is a Machine Learning engineer at PolyAI, a London-based startup looking to use the latest developments in NLP to create a general machine learning platform for deploying spoken dialogue systems. He got his PhD from the University of Sheffield and later he worked as Research Assistant in the Dialogue Systems group, University of Cambridge. His main research interest focuses on increasing the scalability of machine learning based dialogue management, looking for methods to make deep learning and/or reinforcement learning applicable to real world dialogue management tasks. He has published several papers on the topic, two of them nominated to best paper award.

Ivan Vulic (iv250@cam.ac.uk, <https://sites.google.com/site/ivanvulic/>) is a Senior Research Associate in the Language Technology Lab at the University of Cambridge. He holds a PhD from KU Leuven, obtained summa cum laude. Ivan is interested in representation learning, human language understanding, distributional, lexical, and multi-modal semantics in monolingual and multilingual contexts, and transfer learning for enabling cross-lingual NLP applications. He co-lectured a tutorial on monolingual and multilingual topic models and applications at ECIR 2013 and WSDM 2014, a tutorial on word vector space specialisation at EACL 2017, and a tutorial on cross-lingual word representations at EMNLP 2017. He will lecture a course on word vector space specialisation at ESSLLI 2018. He has given invited talks at academia and industry such as Apple Inc., University of Cambridge, UCL, University of Copenhagen, Paris-Saclay, and Bar-Ilan University.

Welcome Reception



Friday, June 1, 2018, 18:00 – 21:00

Hyatt Regency Hotel (Conference Venue)

Empire Ballroom

<https://neworleans.regency.hyatt.com>

Catch up with your colleagues at the **Welcome Reception!** It will be held immediately following the Tutorials, Friday, June 1st, at 18:00, in the Empire Ballroom of the Hyatt Regency Hotel (the conference venue). Refreshments and a light dinner will be provided, and a cash bar will be available.

3

Main Conference: Saturday, June 2

Overview

07:30 – 18:00	Registration	<i>Empire Foyer</i>					
07:30 – 08:45	Breakfast	<i>Empire Foyer</i>					
08:45 – 09:00	Welcome from the Chairs	<i>Empire Ballroom</i>					
09:00 – 10:00	Keynote Address: Charles Yang (sponsored by Toutiao AI Lab)	<i>Empire Ballroom</i>					
10:00 – 10:30	Morning Coffee	<i>Empire Foyer</i>					
10:30 – 11:30	Research Oral 1 <table border="1"> <tr> <td>Information Extraction 1 <i>Empire A</i></td> <td>Phonology, Morphology and Word Segmentation 1 <i>Empire B</i></td> <td>Speech 1 <i>Empire C</i></td> </tr> </table>			Information Extraction 1 <i>Empire A</i>	Phonology, Morphology and Word Segmentation 1 <i>Empire B</i>	Speech 1 <i>Empire C</i>	Industry Oral 1 Poster/Demo 1/2 (10:30-12:00)
Information Extraction 1 <i>Empire A</i>	Phonology, Morphology and Word Segmentation 1 <i>Empire B</i>	Speech 1 <i>Empire C</i>					
	<table border="1"> <tr> <td>Machine Learning 1 <i>Empire A</i></td> <td>Information Extraction 2 <i>Empire B</i></td> <td>Machine Translation 1 <i>Empire C</i></td> </tr> </table>			Machine Learning 1 <i>Empire A</i>	Information Extraction 2 <i>Empire B</i>	Machine Translation 1 <i>Empire C</i>	Machine Learning – Classification <i>Empire D</i> Industry Oral 2 NLP Applications 1 / Discourse and Pragmatics 1 / Generation 1 / Demos
Machine Learning 1 <i>Empire A</i>	Information Extraction 2 <i>Empire B</i>	Machine Translation 1 <i>Empire C</i>					
11:30 – 12:30	<table border="1"> <tr> <td>Machine Learning 1 <i>Empire A</i></td> <td>Information Extraction 2 <i>Empire B</i></td> <td>Machine Translation 1 <i>Empire C</i></td> </tr> </table>			Machine Learning 1 <i>Empire A</i>	Information Extraction 2 <i>Empire B</i>	Machine Translation 1 <i>Empire C</i>	Dialog <i>Empire D</i> Elite Hall B
Machine Learning 1 <i>Empire A</i>	Information Extraction 2 <i>Empire B</i>	Machine Translation 1 <i>Empire C</i>					
12:30 – 14:00	Recruitment Lunch			<i>Elite Hall A</i>			
14:00 – 15:00	Industry Keynote Address: Mari Ostendorf			<i>Empire Ballroom</i>			
15:00 – 15:30	Afternoon Coffee			<i>Empire Foyer</i>			
15:30 – 17:00	Research Oral 3 <table border="1"> <tr> <td>Machine Learning 2 <i>Empire A</i></td> <td>Social Media and Computational Social Science 1 <i>Empire B</i></td> <td>Vision, Robotics and Other Grounding 1 <i>Empire C</i></td> </tr> </table>			Machine Learning 2 <i>Empire A</i>	Social Media and Computational Social Science 1 <i>Empire B</i>	Vision, Robotics and Other Grounding 1 <i>Empire C</i>	Panel Poster/Demo 3 Careers in the Industry Panel Semantics 1 / Sentiment Analysis 1 / Demos
Machine Learning 2 <i>Empire A</i>	Social Media and Computational Social Science 1 <i>Empire B</i>	Vision, Robotics and Other Grounding 1 <i>Empire C</i>					
17:00 – 18:30	Research Oral 4 <table border="1"> <tr> <td>NLP Applications 2 <i>Empire A</i></td> <td>Question Answering 1 <i>Empire B</i></td> <td>SRW highlight papers <i>Empire C</i></td> </tr> </table>			NLP Applications 2 <i>Empire A</i>	Question Answering 1 <i>Empire B</i>	SRW highlight papers <i>Empire C</i>	Industry Oral 4 Machine Translation <i>Empire D</i>
NLP Applications 2 <i>Empire A</i>	Question Answering 1 <i>Empire B</i>	SRW highlight papers <i>Empire C</i>					

Keynote Address: Charles Yang (sponsored by Toutiao AI Lab)

Why 72?

Saturday, June 2, 2018, 9:00–10:00

Chair: Marilyn Walker

Empire Ballroom

Biography: Charles is a Professor of Linguistics, Computer Science, and Psychology at the University of Pennsylvania and directs the Program in Cognitive Science. He has spent a long time to work out the tricks children use to learn languages and is now ready to try them out on machines. His most recent book, *The Price of Linguistic Productivity*, is the winner of the 2017 LSA Leonard Bloomfield award.

Session 1 Overview – Saturday, June 2, 2018

Track A <i>Information Extraction I</i>	Track B <i>Phonology Morphology and Word Segmentation I</i>	Track C <i>Speech I</i>	Track D <i>Industry: Machine Learning – Classification</i>	Track P <i>Posters and Demos (10:30–12:00)</i>
Empire A	Empire B	Empire C	Empire D	Elite Hall B
Label-Aware Double Transfer Learning for Cross-Specialty Medical Named Entity Recognition <i>Z. Wang, Y. Qu, L. Chen, J. Shen, W. Zhang, S. Zhang, Y. Gao, G. Gu, K. Chen, and Y. Yu</i>	A Deep Generative Model of Vowel Formant Typology <i>R. Cotterell and J. Eisner</i>	Parsing Speech: a Neural Approach to Integrating Lexical and Acoustic-Prosodic Information <i>T. Tran, S. Toshniwal, M. Bansal, K. Gimpel, K. Livescu, and M. Ostendorf</i>	Scalable Wide and Deep Learning for Computer Assisted Coding (INDUSTRY) <i>M. Amoia, F. Diehl, J. Gimenez, J. Pinto, R. Schumann, F. Stemmer, P. Vozila, and Y. Zhang</i>	10:30
Neural Fine-Grained Entity Type Classification with Hierarchy-Aware Loss <i>P. Xu and D. Barbosa</i>	Fortification of Neural Morphological Segmentation Models for Polysynthetic Minimal-Resource Languages <i>K. Kann, J. M. Mager Hois, I. V. Meza Ruiz, and H. Schütze</i>	Tied Multitask Learning for Neural Speech Translation <i>A. Anastasopoulos and D. Chiang</i>	Neural Network based Extreme Classification and Similarity Models for Product Matching (INDUSTRY) <i>K. Shah, S. Kopru, and J. D. Ruvini</i>	Discourse and Pragmatics 1 / Generation 1 / NLP Applications 1 / Demos 10:48
Joint Bootstrapping Machines for High Confidence Relation Extraction <i>P. Gupta, B. Roth, and H. Schütze</i>	Improving Character-Based Decoding Using Target-Side Morphological Information for Neural Machine Translation <i>P. Passban, Q. Liu, and A. Way</i>	Please Clap: Modeling Applause in Campaign Speeches <i>J. Gillick and D. Bamman</i>	A Scalable Neural Shortlisting-Reranking Approach for Large-Scale Domain Classification in Natural Language Understanding (INDUSTRY) <i>Y.-B. Kim, D. Kim, J.-K. Kim, and R. Sarikaya</i>	11:06

Parallel Session 1

Oral: Information Extraction 1

Empire A

Chair: Dan Bikel

Label-Aware Double Transfer Learning for Cross-Specialty Medical Named Entity Recognition

Zhenghui Wang, Yanru Qu, Liheng Chen, Jian Shen, Weinan Zhang, Shaodian Zhang, Yimei Gao, Gen Gu, Ken Chen, and Yong Yu

10:30–10:47

We study the problem of named entity recognition (NER) from electronic medical records, which is one of the most fundamental and critical problems for medical text mining. Medical records which are written by clinicians from different specialties usually contain quite different terminologies and writing styles. The difference of specialties and the cost of human annotation makes it particularly difficult to train a universal medical NER system. In this paper, we propose a label-aware double transfer learning framework (La-DTL) for cross-specialty NER, so that a medical NER system designed for one specialty could be conveniently applied to another one with minimal annotation efforts. The transferability is guaranteed by two components: (i) we propose label-aware MMD for feature representation transfer, and (ii) we perform parameter transfer with a theoretical upper bound which is also label aware. We conduct extensive experiments on 12 cross-specialty NER tasks. The experimental results demonstrate that La-DTL provides consistent accuracy improvement over strong baselines. Besides, the promising experimental results on non-medical NER scenarios indicate that La-DTL is potential to be seamlessly adapted to a wide range of NER tasks.

Neural Fine-Grained Entity Type Classification with Hierarchy-Aware Loss

Peng Xu and Denilson Barbosa

10:48–11:05

The task of Fine-grained Entity Type Classification (FETC) consists of assigning types from a hierarchy to entity mentions in text. Existing methods rely on distant supervision and are thus susceptible to noisy labels that can be out-of-context or overly-specific for the training sentence. Previous methods that attempt to address these issues do so with heuristics or with the help of hand-crafted features. Instead, we propose an end-to-end solution with a neural network model that uses a variant of cross-entropy loss function to handle out-of-context labels, and hierarchical loss normalization to cope with overly-specific ones. Also, previous work solve FETC a multi-label classification followed by ad-hoc post-processing. In contrast, our solution is more elegant: we use public word embeddings to train a single-label that jointly learns representations for entity mentions and their context. We show experimentally that our approach is robust against noise and consistently outperforms the state-of-the-art on established benchmarks for the task.

Joint Bootstrapping Machines for High Confidence Relation Extraction

Pankaj Gupta, Benjamin Roth, and Hinrich Schütze

11:06–11:23

Semi-supervised bootstrapping techniques for relationship extraction from text iteratively expand a set of initial seed instances. Due to the lack of labeled data, a key challenge in bootstrapping is semantic drift: if a false positive instance is added during an iteration, then all following iterations are contaminated. We introduce BREX, a new bootstrapping method that protects against such contamination by highly effective confidence assessment. This is achieved by using entity and template seeds jointly (as opposed to just one as in previous work), by expanding entities and templates in parallel and in a mutually constraining fashion in each iteration and by introducing higherquality similarity measures for templates. Experimental results show that BREX achieves an F1 that is 0.13 (0.87 vs. 0.74) better than the state of the art for four relationships.

Oral: Phonology, Morphology and Word Segmentation 1

Empire B

Chair: Barbara Plank

A Deep Generative Model of Vowel Formant Typology

Ryan Cotterell and Jason Eisner

10:30–10:47

What makes some types of languages more probable than others? For instance, we know that almost all spoken languages contain the vowel phoneme /i:/; why should that be? The field of linguistic typology seeks to answer these questions and, thereby, divine the mechanisms that underlie human language. In our work, we tackle the problem of vowel system typology, i.e., we propose a generative probability model of which vowels a language

contains. In contrast to previous work, we work directly with the acoustic information—the first two formant values—rather than modeling discrete sets of symbols from the international phonetic alphabet. We develop a novel generative probability model and report results on over 200 languages.

Fortification of Neural Morphological Segmentation Models for Polysynthetic Minimal-Resource Languages

Katharina Kann, Jesus Manuel Mager Hois, Ivan Vladimir Meza Ruiz, and Hinrich Schütze
10:48–11:05

Morphological segmentation for polysynthetic languages is challenging, because a word may consist of many individual morphemes and training data can be extremely scarce. Since neural sequence-to-sequence (seq2seq) models define the state of the art for morphological segmentation in high-resource settings and for (mostly) European languages, we first show that they also obtain competitive performance for Mexican polysynthetic languages in minimal-resource settings. We then propose two novel multi-task training approaches—one with, one without need for external unlabeled resources—, and two corresponding data augmentation methods, improving over the neural baseline for all languages. Finally, we explore cross-lingual transfer as a third way to fortify our neural model and show that we can train one single multi-lingual model for related languages while maintaining comparable or even improved performance, thus reducing the amount of parameters by close to 75%. We provide our morphological segmentation datasets for Mexicanero, Nahuatl, Wixarika and Yorem Nokki for future research.

Improving Character-Based Decoding Using Target-Side Morphological Information for Neural Machine Translation

Peyman Passban, Qun Liu, and Andy Way 11:06–11:23

Recently, neural machine translation (NMT) has emerged as a powerful alternative to conventional statistical approaches. However, its performance drops considerably in the presence of morphologically rich languages (MRLs). Neural engines usually fail to tackle the large vocabulary and high out-of-vocabulary (OOV) word rate of MRLs. Therefore, it is not suitable to exploit existing word-based models to translate this set of languages. In this paper, we propose an extension to the state-of-the-art model of Chung et al. (2016), which works at the character level and boosts the decoder with target-side morphological information. In our architecture, an additional morphology table is plugged into the model. Each time the decoder samples from a target vocabulary, the table sends auxiliary signals from the most relevant affixes in order to enrich the decoder’s current state and constrain it to provide better predictions. We evaluated our model to translate English into German, Russian, and Turkish as three MRLs and observed significant improvements.

Oral: Speech 1

Empire C

Chair: Dilek Hakkani-Tür

Parsing Speech: a Neural Approach to Integrating Lexical and Acoustic-Prosodic Information

Trang Tran, Shubham Toshniwal, Mohit Bansal, Kevin Gimpel, Karen Livescu, and Mari Ostendorf 10:30–10:47

In conversational speech, the acoustic signal provides cues that help listeners disambiguate difficult parses. For automatically parsing spoken utterances, we introduce a model that integrates transcribed text and acoustic-prosodic features using a convolutional neural network over energy and pitch trajectories coupled with an attention-based recurrent neural network that accepts text and prosodic features. We find that different types of acoustic-prosodic features are individually helpful, and together give statistically significant improvements in parse and disfluency detection F1 scores over a strong text-only baseline. For this study with known sentence boundaries, error analyses show that the main benefit of acoustic-prosodic features is in sentences with disfluencies, attachment decisions are most improved, and transcription errors obscure gains from prosody.

Tied Multitask Learning for Neural Speech Translation

Antonios Anastasopoulos and David Chiang 10:48–11:05

We explore multitask models for neural translation of speech, augmenting them in order to reflect two intuitive notions. First, we introduce a model where the second task decoder receives information from the decoder of the first task, since higher-level intermediate representations should provide useful information. Second, we apply regularization that encourages transitivity and invertibility. We show that the application of these notions on jointly trained models improves performance on the tasks of low-resource speech transcription and translation. It also leads to better performance when using attention information for word discovery over

unsegmented input.

Please Clap: Modeling Applause in Campaign Speeches

Jon Gillick and David Bamman

11:06–11:23

This work examines the rhetorical techniques that speakers employ during political campaigns. We introduce a new corpus of speeches from campaign events in the months leading up to the 2016 U.S. presidential election and develop new models for predicting moments of audience applause. In contrast to existing datasets, we tackle the challenge of working with transcripts that derive from uncorrected closed captioning, using associated audio recordings to automatically extract and align labels for instances of audience applause. In prediction experiments, we find that lexical features carry the most information, but that a variety of features are predictive, including prosody, long-term contextual dependencies, and theoretically motivated features designed to capture rhetorical techniques.

Industry: Machine Learning – Classification

Empire D

Chair: Luke Zettlemoyer

Scalable Wide and Deep Learning for Computer Assisted Coding (INDUSTRY)

Marilisa Amoia, Frank Diehl, Jesus Gimenez, Joel Pinto, Raphael Schumann, Fabian Stemmer, Paul Vozila, and Yi Zhang

10:30–10:47

In recent years the use of electronic medical records has accelerated resulting in large volumes of medical data when a patient visits a healthcare facility. As a first step towards reimbursement healthcare institutions need to associate ICD-10 billing codes to these documents. This is done by trained clinical coders who may use a computer assisted solution for shortlisting of codes. In this work, we present our work to build a machine learning based scalable system for predicting ICD-10 codes from electronic medical records. We address data imbalance issues by implementing two system architectures using convolutional neural networks and logistic regression models. We illustrate the pros and cons of those system designs and show that the best performance can be achieved by leveraging the advantages of both using a system combination approach.

Neural Network based Extreme Classification and Similarity Models for Product Matching (INDUSTRY)

Kashif Shah, Selcuk Koprulu, and Jean David Ruvini

10:48–11:05

Matching a seller listed item to an appropriate product has become a fundamental and one of the most significant step for e-commerce platforms for product based experience. It has a huge impact on making the search effective, search engine optimization, providing product reviews and product price estimation etc. along with many other advantages for a better user experience. As significant and vital it has become, the challenge to tackle the complexity has become huge with the exponential growth of individual and business sellers trading millions of products everyday. We explored two approaches; classification based on shallow neural network and similarity based on deep siamese network. These models outperform the baseline by more than 5% in term of accuracy and are capable of extremely efficient training and inference.

A Scalable Neural Shortlisting-Reranking Approach for Large-Scale Domain Classification in Natural Language Understanding (INDUSTRY)

Young-Bum Kim, Dongchan Kim, Joo-Kyung Kim, and Ruhi Sarikaya

11:06–11:23

Intelligent personal digital assistants (IPDAs), a popular real-life application with spoken language understanding capabilities, can cover potentially thousands of overlapping domains for natural language understanding, and the task of finding the best domain to handle an utterance becomes a challenging problem on a large scale. In this paper, we propose a set of efficient and scalable shortlisting-reranking neural models for effective large-scale domain classification for IPDAs. The shortlisting stage focuses on efficiently trimming all domains down to a list of k-best candidate domains, and the reranking stage performs a list-wise reranking of the initial k-best domains with additional contextual information. We show the effectiveness of our approach with extensive experiments on 1,500 IPDA domains.

Session 2 Overview – Saturday, June 2, 2018

Track A <i>Machine Learning 1</i>	Track B <i>Information Extraction 2</i>	Track C <i>Machine Translation 1</i>	Track D <i>Industry: Dialogos</i>	Track P <i>Posters and Demos (10:30–12:00)</i>
Empire A	Empire B	Empire C	Empire D	Elite Hall B
Zero-shot Sequence Labeling: Transferring Knowledge from Sentences to Tokens <i>M. Rei and A. Søgaard</i>	Learning beyond Datasets: Knowledge Graph Augmented Neural Networks for Natural Language Processing <i>A. K M. S. Basu Roy Chowdhury, and A. Dukkipati</i>	Improving Lexical Choice in Neural Machine Translation <i>T. Nguyen and D. Chiang</i>	What We Need to Learn if We Want to Do and Not Just Talk (INDUSTRY) <i>R. Gangadharaiyah, B. Narayanaswamy, and C. Elkan</i>	11:30
Variable Typing: Assigning Meaning to Variables in Mathematical Text <i>Y. Stathopoulos, S. Baker, M. Rei, and S. Teufel</i>	Comparing Constraints for Taxonomic Organization <i>A. Cocos, M. Apidianaki, and C. Callison-Burch</i>	Universal Neural Machine Translation for Extremely Low Resource Languages <i>J. Gu, H. Hassan, J. Devlin, and V. O. Li</i>	Data Collection for Dialogue System: A Startup Perspective (INDUSTRY) <i>Y. Kang, Y. Zhang, J. K. Kummerfeld, L. Tang, and J. Mars</i>	Discourse and Pragmatics 1 / Generation 1 / NLP Applications 1 / Demos 11:48
Learning Structured Text Representations (TACL) <i>Y. Liu and M. Lapata</i>	Cross-Lingual Syntactic Transfer with Limited Resources (TACL) <i>M. S. Rasooli and M. Collins</i>	Classical Structured Prediction Losses for Sequence to Sequence Learning <i>S. Edunov, M. Ott, M. Auli, D. Grangier, and M. Ranzato</i>	Bootstrapping a Neural Conversational Agent with Dialogue Self-Play, Crowdsourcing, and On-Line Reinforcement Learning (INDUSTRY) <i>P. Shah, D. Hakkani-Tür, B. Liu, and G. Tür</i>	12:06

Parallel Session 2

Oral: Machine Learning 1

Empire A

Chair: Raymond Mooney

Zero-shot Sequence Labeling: Transferring Knowledge from Sentences to Tokens

Marek Rei and Anders Søgaard

11:30–11:47

Can attention- or gradient-based visualization techniques be used to infer token-level labels for binary sequence tagging problems, using networks trained only on sentence-level labels? We construct a neural network architecture based on soft attention, train it as a binary sentence classifier and evaluate against token-level annotation on four different datasets. Inferring token labels from a network provides a method for quantitatively evaluating what the model is learning, along with generating useful feedback in assistance systems. Our results indicate that attention-based methods are able to predict token-level labels more accurately, compared to gradient-based methods, sometimes even rivaling the supervised oracle network.

Variable Typing: Assigning Meaning to Variables in Mathematical Text

Yannos Stathopoulos, Simon Baker, Marek Rei, and Simone Teufel

11:48–12:05

Information about the meaning of mathematical variables in text is useful in NLP/IR tasks such as symbol disambiguation, topic modeling and mathematical information retrieval (MIR). We introduce *variable typing*, the task of assigning one *mathematical type* (multi-word technical terms referring to mathematical concepts) to each variable in a sentence of mathematical text. As part of this work, we also introduce a new annotated data set composed of 33,524 data points extracted from scientific documents published on arXiv. Our intrinsic evaluation demonstrates that our data set is sufficient to successfully train and evaluate current classifiers from three different model architectures. The best performing model is evaluated on an extrinsic task: MIR, by producing a *typed formula index*. Our results show that the best performing MIR models make use of our typed index, compared to a formula index only containing raw symbols, thereby demonstrating the usefulness of variable typing.

Learning Structured Text Representations (TACL)

Yang Liu and Mirella Lapata

12:06–12:23

In this paper, we focus on learning structure-aware document representations from data without recourse to a discourse parser or additional annotations. Drawing inspiration from recent efforts to empower neural networks with a structural bias, we propose a model that can encode a document while automatically inducing rich structural dependencies. Specifically, we embed a differentiable non-projective parsing algorithm into a neural model and use attention mechanisms to incorporate the structural biases. Experimental evaluation across different tasks and datasets shows that the proposed model achieves state-of-the-art results on document modeling tasks while inducing intermediate structures which are both interpretable and meaningful.

Oral: Information Extraction 2

Empire B

Chair: Mihai Surdeanu

Learning beyond Datasets: Knowledge Graph Augmented Neural Networks for Natural Language Processing

Annavaraz K M, Somnath Basu Roy Chowdhury, and Ambedkar Dukkipati

11:30–11:47

Machine Learning has been the quintessential solution for many AI problems, but learning models are heavily dependent on specific training data. Some learning models can be incorporated with prior knowledge using a Bayesian setup, but these learning models do not have the ability to access any organized world knowledge on demand. In this work, we propose to enhance learning models with world knowledge in the form of Knowledge Graph (KG) fact triples for Natural Language Processing (NLP) tasks. Our aim is to develop a deep learning model that can extract relevant prior support facts from knowledge graphs depending on the task using attention mechanism. We introduce a convolution-based model for learning representations of knowledge graph entity and relation clusters in order to reduce the attention space. We show that the proposed method is highly scalable to the amount of prior information that has to be processed and can be applied to any generic NLP task. Using this method we show significant improvement in performance for text classification with 20Newsgroups (News20) & DBpedia datasets, and natural language inference with Stanford Natural

Language Inference (SNLI) dataset. We also demonstrate that a deep learning model can be trained with substantially less amount of labeled training data, when it has access to organized world knowledge in the form of a knowledge base.

Comparing Constraints for Taxonomic Organization

Anne Cocos, Marianna Apidianaki, and Chris Callison-Burch

11:48–12:05

Building a taxonomy from the ground up involves several sub-tasks: selecting terms to include, predicting semantic relations between terms, and selecting a subset of relational instances to keep, given constraints on the taxonomy graph. Methods for this final step – taxonomic organization – vary both in terms of the constraints they impose, and whether they enable discovery of synonymous terms. It is hard to isolate the impact of these factors on the quality of the resulting taxonomy because organization methods are rarely compared directly. In this paper, we present a head-to-head comparison of six taxonomic organization algorithms that vary with respect to their structural and transitivity constraints, and treatment of synonymy. We find that while transitive algorithms out-perform their non-transitive counterparts, the top-performing transitive algorithm is prohibitively slow for taxonomies with as few as 50 entities. We propose a simple modification to a non-transitive optimum branching algorithm to explicitly incorporate synonymy, resulting in a method that is substantially faster than the best transitive algorithm while giving complementary performance.

Cross-Lingual Syntactic Transfer with Limited Resources (TACL)

Mohammad Sadegh Rasooli and Michael Collins

12:06–12:23

We describe a simple but effective method for cross-lingual syntactic transfer of dependency parsers, in the scenario where a large amount of translation data is not available. The method makes use of three steps: 1) a method for deriving cross-lingual word clusters, which can then be used in a multilingual parser; 2) a method for transferring lexical information from a target language to source language treebanks; 3) a method for integrating these steps with the density-driven annotation projection method of Rasooli and Collins (2015). Experiments show improvements over the state-of-the-art in several languages used in previous work, in a setting where the only source of translation data is the Bible, a considerably smaller corpus than the Europarl corpus used in previous work. Results using the Europarl corpus as a source of translation data show additional improvements over the results of Rasooli and Collins (2015). We conclude with results on 38 datasets from the Universal Dependencies corpora.

Oral: Machine Translation 1

Empire C

Chair: Daniel Marcu

Improving Lexical Choice in Neural Machine Translation

Toan Nguyen and David Chiang

11:30–11:47

We explore two solutions to the problem of mistranslating rare words in neural machine translation. First, we argue that the standard output layer, which computes the inner product of a vector representing the context with all possible output word embeddings, rewards frequent words disproportionately, and we propose to fix the norms of both vectors to a constant value. Second, we integrate a simple lexical module which is jointly trained with the rest of the model. We evaluate our approaches on eight language pairs with data sizes ranging from 100k to 8M words, and achieve improvements of up to +4.3 BLEU, surpassing phrase-based translation in nearly all settings.

Universal Neural Machine Translation for Extremely Low Resource Languages

Jiatao Gu, Hany Hassan, Jacob Devlin, and Victor O.K. Li

11:48–12:05

In this paper, we propose a new universal machine translation approach focusing on languages with a limited amount of parallel data. Our proposed approach utilizes a transfer-learning approach to share lexical and sentence level representations across multiple source languages into one target language. The lexical part is shared through a Universal Lexical Representation to support multi-lingual word-level sharing. The sentence-level sharing is represented by a model of experts from all source languages that share the source encoders with all other languages. This enables the low-resource language to utilize the lexical and sentence representations of the higher resource languages. Our approach is able to achieve 23 BLEU on Romanian-English WMT2016 using a tiny parallel corpus of 6k sentences, compared to the 18 BLEU of strong baseline system which uses multi-lingual training and back-translation. Furthermore, we show that the proposed approach can achieve almost 20 BLEU on the same dataset through fine-tuning a pre-trained multi-lingual system in a zero-shot setting.

Classical Structured Prediction Losses for Sequence to Sequence Learning

Sergey Edunov, Myle Ott, Michael Auli, David Grangier, and Marc'Aurelio Ranzato 12:06–12:23

There has been much recent work on training neural attention models at the sequence-level using either reinforcement learning-style methods or by optimizing the beam. In this paper, we survey a range of classical objective functions that have been widely used to train linear models for structured prediction and apply them to neural sequence to sequence models. Our experiments show that these losses can perform surprisingly well by slightly outperforming beam search optimization in a like for like setup. We also report new state of the art results on both IWSLT'14 German-English translation as well as Gigaword abstractive summarization. On the large WMT'14 English-French task, sequence-level training achieves 41.5 BLEU which is on par with the state of the art.

Industry: Dialog

Empire D

Chair: Christine Doran

What We Need to Learn if We Want to Do and Not Just Talk (INDUSTRY)

Rashmi Gangadharaiyah, Balakrishnan Narayanaswamy, and Charles Elkan 11:30–11:47

In task-oriented dialog, agents need to generate both fluent natural language responses and correct external actions like database queries and updates. Our paper makes the first attempt at evaluating state of the art models on a large real world task with human users. We show that methods that achieve state of the art performance on synthetic datasets, perform poorly in real world dialog tasks. We propose a hybrid model, where nearest neighbor is used to generate fluent responses and Seq2Seq type models ensure dialogue coherency and generate accurate external actions. The hybrid model on the customer support data achieves a 78% relative improvement in fluency, and a 200% improvement in accuracy of external calls.

Data Collection for Dialogue System: A Startup Perspective (INDUSTRY)

Yiping Kang, Yunqi Zhang, Jonathan K. Kummerfeld, Lingjia Tang, and Jason Mars 11:48–12:05

Industrial dialogue systems such as Apple Siri and Google Now rely on large scale diverse and robust training data to enable their sophisticated conversation capability. Crowdsourcing provides a scalable and inexpensive way of data collection but collecting high quality data efficiently requires thoughtful orchestration of the crowdsourcing jobs. Prior study of this topic have focused on tasks only in the academia settings with limited scope or only provide intrinsic dataset analysis, lacking indication on how it affects the trained model performance. In this paper, we present a study of crowdsourcing methods for a user intent classification task in our deployed dialogue system. Our task requires classification of 47 possible user intents and contains many intent pairs with subtle differences. We consider different crowdsourcing job types and job prompts and analyze quantitatively the quality of the collected data and the downstream model performance on a test set of real user queries from production logs. Our observation provides insights into designing efficient crowdsourcing jobs and provide recommendations for future dialogue system data collection process.

Bootstrapping a Neural Conversational Agent with Dialogue Self-Play, Crowdsourcing and On-Line Reinforcement Learning (INDUSTRY)

Pararth Shah, Dilek Hakkani-Tür, Bing Liu, and Gokhan Tür 12:06–12:23

End-to-end neural models show great promise towards building conversational agents that are trained from data and on-line experience using supervised and reinforcement learning. However, these models require a large corpus of dialogues to learn effectively. For goal-oriented dialogues, such datasets are expensive to collect and annotate, since each task involves a separate schema and database of entities. Further, the Wizard-of-Oz approach commonly used for dialogue collection does not provide sufficient coverage of salient dialogue flows, which is critical for guaranteeing an acceptable task completion rate in consumer-facing conversational agents. In this paper, we study a recently proposed approach for building an agent for arbitrary tasks by combining dialogue self-play and crowd-sourcing to generate fully-annotated dialogues with diverse and natural utterances. We discuss the advantages of this approach for industry applications of conversational agents, wherein an agent can be rapidly bootstrapped to deploy in front of users and further optimized via interactive learning from actual users of the system.

Posters and Demos: Sessions 1/2

Posters: Discourse and Pragmatics 1

Time: 10:30–12:00

Location: Elite Hall B

Chair: Laura Perez-Beltrachini

Attentive Interaction Model: Modeling Changes in View in Argumentation

Yohan Jo, Shivani Poddar, Byungssoo Jeon, Qinlan Shen, Carolyn Rose, and Graham Neubig

We present a neural architecture for modeling argumentative dialogue that explicitly models the interplay between an Opinion Holder's (OH's) reasoning and a challenger's argument, with the goal of predicting if the argument successfully changes the OH's view. The model has two components: (1) vulnerable region detection, an attention model that identifies parts of the OH's reasoning that are amenable to change, and (2) interaction encoding, which identifies the relationship between the content of the OH's reasoning and that of the challenger's argument. Based on evaluation on discussions from the Change My View forum on Reddit, the two components work together to predict an OH's change in view, outperforming several baselines. A posthoc analysis suggests that sentences picked out by the attention model are addressed more frequently by successful arguments than by unsuccessful ones.

Automatic Focus Annotation: Bringing Formal Pragmatics Alive in Analyzing the Information Structure of Authentic Data

Ramon Ziai and Detmar Meurers

Analyzing language in context, both from a theoretical and from a computational perspective, is receiving increased interest. Complementing the research in linguistics on discourse and information structure, in computational linguistics identifying discourse concepts was also shown to improve the performance of certain applications, for example, Short Answer Assessment systems (Ziai and Meurers, 2014). Building on the research that established detailed annotation guidelines for manual annotation of information structural concepts for written (Dipper et al., 2007; Ziai and Meurers, 2014) and spoken language data (Calhoun et al., 2010), this paper presents the first approach automating the analysis of focus in authentic written data. Our classification approach combines a range of lexical, syntactic, and semantic features to achieve an accuracy of 78.1% for identifying focus.

Dear Sir or Madam, May I Introduce the GYAFC Dataset: Corpus, Benchmarks and Metrics for Formality Style Transfer

Sudha Rao and Joel Tetreault

Style transfer is the task of automatically transforming a piece of text in one particular style into another. A major barrier to progress in this field has been a lack of training and evaluation datasets, as well as benchmarks and automatic metrics. In this work, we create the largest corpus for a particular stylistic transfer (formality) and show that techniques from the machine translation community can serve as strong baselines for future work. We also discuss challenges of using automatic metrics.

Enhanced Word Representations for Bridging Anaphora Resolution

Yufang Hou

Most current models of word representations (e.g., GloVe) have successfully captured fine-grained semantics. However, semantic similarity exhibited in these word embeddings is not suitable for resolving bridging anaphora, which requires the knowledge of associative similarity (i.e., relatedness) instead of semantic similarity information between synonyms or hypernyms. We create word embeddings (embeddings_PP) to capture such relatedness by exploring the syntactic structure of noun phrases. We demonstrate that using embeddings_PP alone achieves around 30% of accuracy for bridging anaphora resolution on the ISNotes corpus. Furthermore, we achieve a substantial gain over the state-of-the-art system (Hou et al., 2013b) for bridging antecedent selection.

Gender Bias in Coreference Resolution

Rachel Rudinger, Jason Naradowsky, Brian Leonard, and Benjamin Van Durme

We present an empirical study of gender bias in coreference resolution systems. We first introduce a novel, Winograd schema-style set of minimal pair sentences that differ only by pronoun gender. With these “Winogender schemas,” we evaluate and confirm systematic gender bias in three publicly-available coreference resolution systems, and correlate this bias with real-world and textual gender statistics.

Gender Bias in Coreference Resolution: Evaluation and Debiasing Methods*Jieyu Zhao, Tianlu Wang, Mark Yatskar, Vicente Ordonez, and Kai-Wei Chang*

In this paper, we introduce a new benchmark for co-reference resolution focused on gender bias, WinoBias. Our corpus contains Winograd-schema style sentences with entities corresponding to people referred by their occupation (e.g. the nurse, the doctor, the carpenter). We demonstrate that a rule-based, a feature-rich, and a neural coreference system all link gendered pronouns to pro-stereotypical entities with higher accuracy than anti-stereotypical entities, by an average difference of 21.1 in F1 score. Finally, we demonstrate a data-augmentation approach that, in combination with existing word-embedding debiasing techniques, removes the bias demonstrated by these systems in WinoBias without significantly affecting their performance on existing datasets.

Improving Implicit Discourse Relation Classification by Modeling Inter-dependencies of Discourse Units in a Paragraph*Zeyu Dai and Ruihong Huang*

We argue that semantic meanings of a sentence or clause can not be interpreted independently from the rest of a paragraph, or independently from all discourse relations and the overall paragraph-level discourse structure. With the goal of improving implicit discourse relation classification, we introduce a paragraph-level neural networks that model inter-dependencies between discourse units as well as discourse relation continuity and patterns, and predict a sequence of discourse relations in a paragraph. Experimental results show that our model outperforms the previous state-of-the-art systems on the benchmark corpus of PDTB.

Integrating Stance Detection and Fact Checking in a Unified Corpus*Ramy Baly, Mitra Mohtarami, James Glass, Lluís Marquez, Alessandro Moschitti, and Preslav Nakov*

A reasonable approach for fact checking a claim involves retrieving potentially relevant documents from different sources (e.g., news websites, social media, etc.), determining the stance of each document with respect to the claim, and finally making a prediction about the claim's factuality by aggregating the strength of the stances, while taking the reliability of the source into account. Moreover, a fact checking system should be able to explain its decision by providing relevant extracts (rationales) from the documents. Yet, this setup is not directly supported by existing datasets, which treat fact checking, document retrieval, source credibility, stance detection and rationale extraction as independent tasks. In this paper, we support the interdependencies between these tasks as annotations in the same corpus. We implement this setup on an Arabic fact checking corpus, the first of its kind.

Is Something Better than Nothing? Automatically Predicting Stance-based Arguments Using Deep Learning and Small Labelled Dataset*Pavithra Rajendran, Danushka Bollegala, and Simon Parsons*

Online reviews have become a popular portal among customers making decisions about purchasing products. A number of corpora of reviews have been widely investigated in NLP in general, and, in particular, in argument mining. This is a subset of NLP that deals with extracting arguments and the relations among them from user-based content. A major problem faced by argument mining research is the lack of human-annotated data. In this paper, we investigate the use of weakly supervised and semi-supervised methods for automatically annotating data, and thus providing large annotated datasets. We do this by building on previous work that explores the classification of opinions present in reviews based whether the stance is expressed explicitly or implicitly. In the work described here, we automatically annotate stance as implicit or explicit and our results show that the datasets we generate, although noisy, can be used to learn better models for implicit/explicit opinion classification.

Multi-Task Learning for Argumentation Mining in Low-Resource Settings*Claudia Schulz, Steffen Eger, Johannes Daxenberger, Tobias Kahse, and Iryna Gurevych*

We investigate whether and where multi-task learning (MTL) can improve performance on NLP problems related to argumentation mining (AM), in particular argument component identification. Our results show that MTL performs particularly well (and better than single-task learning) when little training data is available for the main task, a common scenario in AM. Our findings challenge previous assumptions that conceptualizations across AM datasets are divergent and that MTL is difficult for semantic or higher-level tasks.

Neural Models for Reasoning over Multiple Mentions Using Coreference*Bhuwan Dhingra, Qiao Jin, Zhilin Yang, William Cohen, and Ruslan Salakhutdinov*

Many problems in NLP require aggregating information from multiple mentions of the same entity which may be far apart in the text. Existing Recurrent Neural Network (RNN) layers are biased towards short-

term dependencies and hence not suited to such tasks. We present a recurrent layer which is instead biased towards coreferent dependencies. The layer uses coreference annotations extracted from an external system to connect entity mentions belonging to the same cluster. Incorporating this layer into a state-of-the-art reading comprehension model improves performance on three datasets – Wikihop, LAMBADA and the bAbi AI tasks – with large gains when training data is scarce.

Alignment, Acceptance, and Rejection of Group Identities in Online Political Discourse (SRW)

Hagyeong Shin and Gabriel Doyle

Conversation is a joint social process, with participants cooperating to exchange information. This process is helped along through linguistic alignment: participants' adoption of each other's word use. This alignment is robust, appearing many settings, and is nearly always positive. We create an alignment model for examining alignment in Twitter conversations across antagonistic groups. This model finds that some word categories, specifically pronouns used to establish group identity and common ground, are negatively aligned. This negative alignment is observed despite other categories, which are less related to the group dynamics, showing the standard positive alignment. This suggests that alignment is strongly biased toward cooperative alignment, but that different linguistic features can show substantially different behaviors.

Posters: Generation 1

Time: 10:30–12:00

Location: Elite Hall B

Chair: Laura Perez-Beltrachini

A Deep Ensemble Model with Slot Alignment for Sequence-to-Sequence Natural Language Generation

Juraj Juraska, Panagiotis Karagiannis, Kevin Bowden, and Marilyn Walker

Natural language generation lies at the core of generative dialogue systems and conversational agents. We describe an ensemble neural language generator, and present several novel methods for data representation and augmentation that yield improved results in our model. We test the model on three datasets in the restaurant, TV and laptop domains, and report both objective and subjective evaluations of our best model. Using a range of automatic metrics, as well as human evaluators, we show that our approach achieves better results than state-of-the-art models on the same datasets.

A Melody-Conditioned Lyrics Language Model

Kento Watanabe, Yuichiro Matsubayashi, Satoru Fukayama, Masataka Goto, Kentaro Inui, and Tomoyasu Nakano

This paper presents a novel, data-driven language model that produces entire lyrics for a given input melody. Previously proposed models for lyrics generation suffer from the inability of capturing the relationship between lyrics and melody partly due to the unavailability of lyrics-melody aligned data. In this study, we first propose a new practical method for creating a large collection of lyrics-melody aligned data and then create a collection of 1,000 lyrics-melody pairs augmented with precise syllable-note alignments and word/sentence/paragraph boundaries. We then provide a quantitative analysis of the correlation between word/sentence/paragraph boundaries in lyrics and melodies. We then propose an RNN-based lyrics language model conditioned on a featurized melody. Experimental results show that the proposed model generates fluent lyrics while maintaining the compatibility between boundaries of lyrics and melody structures.

Automatic Dialogue Generation with Expressed Emotions

Chenyang Huang, Osmar Zaiane, Amine Trabelsi, and Nouha Dziri

Despite myriad efforts in the literature designing neural dialogue generation systems in recent years, very few consider putting restrictions on the response itself. They learn from collections of past responses and generate one based on a given utterance without considering, speech act, desired style or emotion to be expressed. In this research, we address the problem of forcing the dialogue generation to express emotion. We present three models that either concatenate the desired emotion with the source input during the learning, or push the emotion in the decoder. The results, evaluated with an emotion tagger, are encouraging with all three models, but present better outcome and promise with our model that adds the emotion vector in the decoder.

Discourse-Aware Neural Rewards for Coherent Text Generation*Antoine Bosselut, Asli Celikyilmaz, Xiaodong He, Jianfeng Gao, Po-Sen Huang, and Yejin Choi*

In this paper, we investigate the use of discourse-aware rewards with reinforcement learning to guide a model to generate long, coherent text. In particular, we propose to learn neural rewards to model cross-sentence ordering as a means to approximate desired discourse structure. Empirical results demonstrate that a generator trained with the learned reward produces more coherent and less repetitive text than models trained with cross-entropy or with reinforcement learning with commonly used scores as rewards.

Guiding Generation for Abstractive Text Summarization Based on Key Information Guide Network*Chenliang Li, Weiran Xu, Si Li, and Sheng Gao*

Neural network models, based on the attentional encoder-decoder model, have good capability in abstractive text summarization. However, these models are hard to be controlled in the process of generation, which leads to a lack of key information. We propose a guiding generation model that combines the extractive method and the abstractive method. Firstly, we obtain keywords from the text by a extractive model. Then, we introduce a Key Information Guide Network (KIGN), which encodes the keywords to the key information representation, to guide the process of generation. In addition, we use a prediction-guide mechanism, which can obtain the long-term value for future decoding, to further guide the summary generation. We evaluate our model on the CNN/Daily Mail dataset. The experimental results show that our model leads to significant improvements.

Natural Answer Generation with Heterogeneous Memory*Yao Fu and Yansong Feng*

Memory augmented encoder-decoder framework has achieved promising progress for natural language generation tasks. Such frameworks enable a decoder to retrieve from a memory during generation. However, less research has been done to take care of the memory contents from different sources, which are often of heterogeneous formats. In this work, we propose a novel attention mechanism to encourage the decoder to actively interact with the memory by taking its heterogeneity into account. Our solution attends across the generated history and memory to explicitly avoid repetition, and introduce related knowledge to enrich our generated sentences. Experiments on the answer sentence generation task show that our method can effectively explore heterogeneous memory to produce readable and meaningful answer sentences while maintaining high coverage for given answer information.

Natural Language Generation by Hierarchical Decoding with Linguistic Patterns*Shang-Yu Su, Kai-Ling Lo, Yi Ting Yeh, and Yun-Nung Chen*

Natural language generation (NLG) is a critical component in spoken dialogue systems. Classic NLG can be divided into two phases: (1) sentence planning: deciding on the overall sentence structure, (2) surface realization: determining specific word forms and flattening the sentence structure into a string. Many simple NLG models are based on recurrent neural networks (RNN) and sequence-to-sequence (seq2seq) model, which basically contains a encoder-decoder structure; these NLG models generate sentences from scratch by jointly optimizing sentence planning and surface realization using a simple cross entropy loss training criterion. However, the simple encoder-decoder architecture usually suffers from generating complex and long sentences, because the decoder has to learn all grammar and diction knowledge. This paper introduces a hierarchical decoding NLG model based on linguistic patterns in different levels, and shows that the proposed method outperforms the traditional one with a smaller model size. Furthermore, the design of the hierarchical decoding is flexible and easily-extendible in various NLG systems.

Neural Poetry Translation*Marjan Ghazvininejad, Yejin Choi, and Kevin Knight*

We present the first neural poetry translation system. Unlike previous works that often fail to produce any translation for fixed rhyme and rhythm patterns, our system always translates a source text to an English poem. Human evaluation of the translations ranks the quality as acceptable 78.2% of the time.

Query and Output: Generating Words by Querying Distributed Word Representations for Paraphrase Generation*Shuming Ma, Xu Sun, Wei Li, Sujian Li, Wenjie Li, and Xuancheng Ren*

Most recent approaches use the sequence-to-sequence model for paraphrase generation. The existing sequence-to-sequence model tends to memorize the words and the patterns in the training dataset instead of learning the meaning of the words. Therefore, the generated sentences are often grammatically correct but semantically improper. In this work, we introduce a novel model based on the encoder-decoder framework, called Word Embedding Attention Network (WEAN). Our proposed model generates the words by querying distributed

word representations (i.e. neural word embeddings), hoping to capturing the meaning of the according words. Following previous work, we evaluate our model on two paraphrase-oriented tasks, namely text simplification and short text abstractive summarization. Experimental results show that our model outperforms the sequence-to-sequence baseline by the BLEU score of 6.3 and 5.5 on two English text simplification datasets, and the ROUGE-2 F1 score of 5.7 on a Chinese summarization dataset. Moreover, our model achieves state-of-the-art performances on these three benchmark datasets.

RankME: Reliable Human Ratings for Natural Language Generation

Jekaterina Novikova, Ondřej Dušek, and Verena Rieser

Human evaluation for natural language generation (NLG) often suffers from inconsistent user ratings. While previous research tends to attribute this problem to individual user preferences, we show that the quality of human judgements can also be improved by experimental design. We present a novel rank-based magnitude estimation method (RankME), which combines the use of continuous scales and relative assessments. We show that RankME significantly improves the reliability and consistency of human ratings compared to traditional evaluation methods. In addition, we show that it is possible to evaluate NLG systems according to multiple, distinct criteria, which is important for error analysis. Finally, we demonstrate that RankME, in combination with Bayesian estimation of system quality, is a cost-effective alternative for ranking multiple NLG systems.

Sentence Simplification with Memory-Augmented Neural Networks

Tu Vu, Baotian Hu, Tsendsuren Munkhdalai, and Hong Yu

Sentence simplification aims to simplify the content and structure of complex sentences, and thus make them easier to interpret for human readers, and easier to process for downstream NLP applications. Recent advances in neural machine translation have paved the way for novel approaches to the task. In this paper, we adapt an architecture with augmented memory capacities called Neural Semantic Encoders (Munkhdalai and Yu, 2017) for sentence simplification. Our experiments demonstrate the effectiveness of our approach on different simplification datasets, both in terms of automatic evaluation measures and human judgments.

Simplification Using Paraphrases and Context-Based Lexical Substitution

Reno Kriz, Eleni Miltzakaki, Marianna Apidianaki, and Chris Callison-Burch

Lexical simplification involves identifying complex words or phrases that need to be simplified, and recommending simpler meaning-preserving substitutes that can be more easily understood. We propose a complex word identification (CWI) model that exploits both lexical and contextual features, and a simplification mechanism which relies on a word-embedding lexical substitution model to replace the detected complex words with simpler paraphrases. We compare our CWI and lexical simplification models to several baselines, and evaluate the performance of our simplification system against human judgments. The results show that our models are able to detect complex words with higher accuracy than other commonly used methods, and propose good simplification substitutes in context. They also highlight the limited contribution of context features for CWI, which nonetheless improve simplification compared to context-unaware models.

Zero-Shot Question Generation from Knowledge Graphs for Unseen Predicates and Entity Types

Hady Elsayar, Christophe Gravier, and Frédérique Laforest

We present a neural model for question generation from knowledge graphs triples in a “Zero-shot” setup, that is generating questions for predicate, subject types or object types that were not seen at training time. Our model leverages triples occurrences in the natural language corpus in a encoder-decoder architecture, paired with an original part-of-speech copy action mechanism to generate questions. Benchmark and human evaluation show that our model outperforms state-of-the-art on this task.

Quality Estimation for Automatically Generated Titles of eCommerce Browse Pages (INDUSTRY)

Nicola Ueffing, José G. C. de Souza, and Gregor Leusch

At eBay, we are automatically generating a large amount of natural language titles for eCommerce browse pages using machine translation (MT) technology. While automatic approaches can generate millions of titles very fast, they are prone to errors. We therefore develop quality estimation (QE) methods which can automatically detect titles with low quality in order to prevent them from going live. In this paper, we present different approaches: The first one is a Random Forest (RF) model that explores hand-crafted, robust features, which are a mix of established features commonly used in Machine Translation Quality Estimation (MTQE) and new features developed specifically for our task. The second model is based on Siamese Networks (SNs) which embed the metadata input sequence and the generated title in the same space and do not require hand-crafted

features at all. We thoroughly evaluate and compare those approaches on in-house data. While the RF models are competitive for scenarios with smaller amounts of training data and somewhat more robust, they are clearly outperformed by the SN models when the amount of training data is larger.

Posters: NLP Applications 1

Time: 10:30–12:00

Location: Elite Hall B

Chair: *Laura Perez-Beltrachini*

A Corpus of Non-Native Written English Annotated for Metaphor

Beata Beigman Klebanov, Chee Wee (Ben) Leong, and Michael Flor

We present a corpus of 240 argumentative essays written by non-native speakers of English annotated for metaphor. The corpus is made publicly available. We provide benchmark performance of state-of-the-art systems on this new corpus, and explore the relationship between writing proficiency and metaphor use.

A Simple and Effective Approach to the Story Cloze Test

Siddarth Srinivasan, Richa Arora, and Mark Riedl

In the Story Cloze Test, a system is presented with a 4-sentence prompt to a story, and must determine which one of two potential endings is the ‘right’ ending to the story. Previous work has shown that ignoring the training set and training a model on the validation set can achieve high accuracy on this task due to stylistic differences between the story endings in the training set and validation and test sets. Following this approach, we present a simpler fully-neural approach to the Story Cloze Test using skip-thought embeddings of the stories in a feed-forward network that achieves close to state-of-the-art performance on this task without any feature engineering. We also find that considering just the last sentence of the prompt instead of the whole prompt yields higher accuracy with our approach.

An Annotated Corpus for Machine Reading of Instructions in Wet Lab Protocols

Chaitanya Kulkarni, Wei Xu, Alan Ritter, and Raghu Machiraju

We describe an effort to annotate a corpus of natural language instructions consisting of 622 wet lab protocols to facilitate automatic or semi-automatic conversion of protocols into a machine-readable format and benefit biological research. Experimental results demonstrate the utility of our corpus for developing machine learning approaches to shallow semantic parsing of instructional texts. We make our annotated Wet Lab Protocol Corpus available to the research community.

Annotation Artifacts in Natural Language Inference Data

Suchin Gururangan, Swabha Swayamdipta, Omer Levy, Roy Schwartz, Samuel Bowman, and Noah A. Smith

Large-scale datasets for natural language inference are created by presenting crowd workers with a sentence (premise), and asking them to generate three new sentences (hypotheses) that it entails, contradicts, or is logically neutral with respect to. We show that, in a significant portion of such data, this protocol leaves clues that make it possible to identify the label by looking only at the hypothesis, without observing the premise. Specifically, we show that a simple text categorization model can correctly classify the hypothesis alone in about 67% of SNLI (Bowman et. al, 2015) and 53% of MultiNLI (Williams et. al, 2017). Our analysis reveals that specific linguistic phenomena such as negation and vagueness are highly correlated with certain inference classes. Our findings suggest that the success of natural language inference models to date has been overestimated, and that the task remains a hard open problem.

Automated Essay Scoring in the Presence of Biased Ratings

Evelin Amorim, Marcia Cançado, and Adriano Vélosio

Studies in Social Sciences have revealed that when people evaluate someone else, their evaluations often reflect their biases. As a result, rater bias may introduce highly subjective factors that make their evaluations inaccurate. This may affect automated essay scoring models in many ways, as these models are typically designed to model (potentially biased) essay raters. While there is sizeable literature on rater effects in general settings, it remains unknown how rater bias affects automated essay scoring. To this end, we present a new annotated corpus containing essays and their respective scores. Different from existing corpora, our corpus also contains comments provided by the raters in order to ground their scores. We present features to quantify rater bias based on their comments, and we found that rater bias plays an important role in automated essay scoring. We investigated the extent to which rater bias affects models based on hand-crafted features. Finally, we propose to rectify the training set by removing essays associated with potentially biased scores while

learning the scoring model.

Content-Based Citation Recommendation

Chandra Bhagavatula, Sergey Feldman, Russell Power, and Waleed Ammar

We present a content-based method for recommending citations in an academic paper draft. We embed a given query document into a vector space, then use its nearest neighbors as candidates, and rerank the candidates using a discriminative model trained to distinguish between observed and unobserved citations. Unlike previous work, our method does not require metadata such as author names which can be missing, e.g., during the peer review process. Without using metadata, our method outperforms the best reported results on PubMed and DBLP datasets with relative improvements of over 18% in F120 and over 22% in MRR. We show empirically that, although adding metadata improves the performance on standard metrics, it favors self-citations which are less useful in a citation recommendation setup. We release an online portal for citation recommendation based on our method, (URL: <http://bit.ly/citeDemo>) and a new dataset OpenCorpus of 7 million research articles to facilitate future research on this task.

Humor Recognition Using Deep Learning

Peng-Yu Chen and Von-Wun Soo

Humor is an essential but most fascinating element in personal communication. How to build computational models to discover the structures of humor, recognize humor and even generate humor remains a challenge and there have been yet few attempts on it. In this paper, we construct and collect four datasets with distinct joke types in both English and Chinese and conduct learning experiments on humor recognition. We implement a Convolutional Neural Network (CNN) with extensive filter size, number and Highway Networks to increase the depth of networks. Results show that our model outperforms in recognition of different types of humor with benchmarks collected in both English and Chinese languages on accuracy, precision, and recall in comparison to previous works.

Leveraging Intra-User and Inter-User Representation Learning for Automated Hate Speech Detection

Jing Qian, Mai ElSherief, Elizabeth Belding, and William Yang Wang

Hate speech detection is a critical, yet challenging problem in Natural Language Processing (NLP). Despite the existence of numerous studies dedicated to the development of NLP hate speech detection approaches, the accuracy is still poor. The central problem is that social media posts are short and noisy, and most existing hate speech detection solutions take each post as an isolated input instance, which is likely to yield high false positive and negative rates. In this paper, we radically improve automated hate speech detection by presenting a novel model that leverages intra-user and inter-user representation learning for robust hate speech detection on Twitter. In addition to the target Tweet, we collect and analyze the user's historical posts to model intra-user Tweet representations. To suppress the noise in a single Tweet, we also model the similar Tweets posted by all other users with reinforced inter-user representation learning techniques. Experimentally, we show that leveraging these two representations can significantly improve the f-score of a strong bidirectional LSTM baseline model by 10.1%.

Looking Beyond the Surface: A Challenge Set for Reading Comprehension over Multiple Sentences

Daniel Khashabi, Snigdha Chaturvedi, Michael Roth, Shyam Upadhyay, and Dan Roth

We present a reading comprehension challenge in which questions can only be answered by taking into account information from multiple sentences. We solicit and verify questions and answers for this challenge through a 4-step crowdsourcing experiment. Our challenge dataset contains 6,500+ questions for 1000+ paragraphs across 7 different domains (elementary school science, news, travel guides, fiction stories, etc) bringing in linguistic diversity to the texts and to the questions wordings. On a subset of our dataset, we found human solvers to achieve an F1-score of 88.1%. We analyze a range of baselines, including a recent state-of-art reading comprehension system, and demonstrate the difficulty of this challenge, despite a high human performance. The dataset is the first to study multi-sentence inference at scale, with an open-ended set of question types that requires reasoning skills.

Neural Automated Essay Scoring and Coherence Modeling for Adversarially Crafted Input

Youmna Farag, Helen Yannakoudakis, and Ted Briscoe

We demonstrate that current state-of-the-art approaches to Automated Essay Scoring (AES) are not well-suited to capturing adversarially crafted input of grammatical but incoherent sequences of sentences. We develop a neural model of local coherence that can effectively learn connectedness features between sentences,

and propose a framework for integrating and jointly training the local coherence model with a state-of-the-art AES model. We evaluate our approach against a number of baselines and experimentally demonstrate its effectiveness on both the AES task and the task of flagging adversarial input, further contributing to the development of an approach that strengthens the validity of neural essay scoring models.

QuickEdit: Editing Text & Translations by Crossing Words Out

David Grangier and Michael Auli

We propose a framework for computer-assisted text editing. It applies to translation post-editing and to paraphrasing. Our proposal relies on very simple interactions: a human editor modifies a sentence by marking tokens they would like the system to change. Our model then generates a new sentence which reformulates the initial sentence by avoiding marked words. The approach builds upon neural sequence-to-sequence modeling and introduces a neural network which takes as input a sentence along with change markers. Our model is trained on translation bitext by simulating post-edits. We demonstrate the advantage of our approach for translation post-editing through simulated post-edits. We also evaluate our model for paraphrasing through a user study.

Reference-less Measure of Faithfulness for Grammatical Error Correction

Leshem Choshen and Omri Abend

We propose USim, a semantic measure for Grammatical Error Correction (that measures the semantic faithfulness of the output to the source, thereby complementing existing reference-less measures (RLMs) for measuring the output's grammaticality. USim operates by comparing the semantic symbolic structure of the source and the correction, without relying on manually-curated references. Our experiments establish the validity of USim, by showing that the semantic structures can be consistently applied to ungrammatical text, that valid corrections obtain a high USIM similarity score to the source, and that invalid corrections obtain a lower score.

Tempo-Lexical Context Driven Word Embedding for Cross-Session Search Task Extraction

Procheta Sen, Debasis Ganguly, and Gareth Jones

Task extraction is the process of identifying search intents over a set of queries potentially spanning multiple search sessions. Most existing research on task extraction has focused on identifying tasks within a single session, where the notion of a session is defined by a fixed length time window. By contrast, in this work we seek to identify tasks that span across multiple sessions. To identify tasks, we conduct a global analysis of a query log in its entirety without restricting analysis to individual temporal windows. To capture inherent task semantics, we represent queries as vectors in an abstract space. We learn the embedding of query words in this space by leveraging the temporal and lexical contexts of queries. Embedded query vectors are then clustered into tasks. Experiments demonstrate that task extraction effectiveness is improved significantly with our proposed method of query vector embedding in comparison to existing approaches that make use of documents retrieved from a collection to estimate semantic similarities between queries.

Atypical Inputs in Educational Applications (INDUSTRY)

Su-Youn Yoon, Aoife Cahill, Anastassia Loukina, Klaus Zechner, Brian Riordan, and Nitin Madnani

In large-scale educational assessments, the use of automated scoring has recently become quite common. While the majority of student responses can be processed and scored without difficulty, there are a small number of responses that have atypical characteristics that make it difficult for an automated scoring system to assign a correct score. We describe a pipeline that detects and processes these kinds of responses at run-time. We present the most frequent kinds of what are called non-scorable responses along with effective filtering models based on various NLP and speech processing technologies. We give an overview of two operational automated scoring systems—one for essay scoring and one for speech scoring—and describe the filtering models they use. Finally, we present an evaluation and analysis of filtering models used for spoken responses in an assessment of language proficiency.

Using Aspect Extraction Approaches to Generate Review Summaries and User Profiles (INDUSTRY)

Christopher Mitcheltree, Veronica Wharton, and Avneesh Saluja

Reviews of products or services on Internet marketplace websites contain a rich amount of information. Users often wish to survey reviews or review snippets from the perspective of a certain aspect, which has resulted in a large body of work on aspect identification and extraction from such corpora. In this work, we evaluate a newly-proposed neural model for aspect extraction on two practical tasks. The first is to extract canonical sentences of various aspects from reviews, and is judged by human evaluators against alternatives. A k -means baseline does remarkably well in this setting. The second experiment focuses on the suitability of the

recovered aspect distributions to represent users by the reviews they have written. Through a set of review reranking experiments, we find that aspect-based profiles can largely capture notions of user preferences, by showing that divergent users generate markedly different review rankings.

SystemT: Declarative Text Understanding for Enterprise (INDUSTRY)

Laura Chiticariu, Marina Danilevsky, Yunyao Li, Frederick Reiss, and Huaiyu Zhu

The rise of enterprise applications over unstructured and semi-structured documents poses new challenges to text understanding systems across multiple dimensions. We present SystemT, a declarative text understanding system that addresses these challenges and has been deployed in a wide range of enterprise applications. We highlight the design considerations and decisions behind SystemT in addressing the needs of the enterprise setting. We also summarize the impact of SystemT on business and education.

Construction of the Literature Graph in Semantic Scholar (INDUSTRY)

Waleed Ammar, Dirk Groeneveld, Chandra Bhagavatula, Iz Beltagy, Miles Crawford, Doug Downey, Jason Dunkelberger, Ahmed Elgohary, Sergey Feldman, Vu Ha, Rodney Kinney, Sebastian Kohlmeier, Kyle Lo, Tyler Murray, Hsu-Han Ooi, Matthew Peters, Joanna Power, Sam Skjonsberg, Lucy Wang, Chris Willhelm, Zheng Yuan, Madeleine van Zuylen, and Oren Etzioni

We describe a deployed scalable system for organizing published scientific literature into a heterogeneous graph to facilitate algorithmic manipulation and discovery. The resulting literature graph consists of more than 280M nodes, representing papers, authors, entities and various interactions between them (e.g., authorships, citations, entity mentions). We reduce literature graph construction into familiar NLP tasks (e.g., entity extraction and linking), point out research challenges due to differences from standard formulations of these tasks, and report empirical results for each task. The methods described in this paper are used to enable semantic features in www.semanticscholar.org.

Demos

Time: 10:30–12:00

Location: Elite Hall B

NLP Lean Programming Framework: Developing NLP Applications More Effectively

Marc Schreiber, Bodo Kraft, and Albert Zündorf

This paper presents NLP Lean Programming framework (NLPf), a new framework for creating custom Natural Language Processing (NLP) models and pipelines by utilizing common software development build systems. This approach allows developers to train and integrate domain-specific NLP pipelines into their applications seamlessly. Additionally, NLPf provides an annotation tool which improves the annotation process significantly by providing a well-designed GUI and sophisticated way of using input devices. Due to NLPf's properties developers and domain experts are able to build domain-specific NLP application more effectively. Project page: <https://gitlab.com/schrieveslaach/NLPf> Video Tutorial: <https://www.youtube.com/watch?v=44UJspVebTA> (Demonstration starts at 11:40 min) This paper is related to: - Interfaces and resources to support linguistic annotation - Software architectures and reusable components - Software tools for evaluation or error analysis

Pay-Per-Request Deployment of Neural Network Models Using Serverless Architectures

Zhucheng Tu, Mengping Li, and Jimmy Lin

We demonstrate the serverless deployment of neural networks for model inferencing in NLP applications using Amazon's Lambda service for feedforward evaluation and DynamoDB for storing word embeddings. Our architecture realizes a pay-per-request pricing model, requiring zero ongoing costs for maintaining server instances. All virtual machine management is handled behind the scenes by the cloud provider without any direct developer intervention. We describe a number of techniques that allow efficient use of serverless resources, and evaluations confirm that our design is both scalable and inexpensive.

An automated medical scribe for documenting clinical encounters

Gregory Finley, Erik Edwards, Amanda Robinson, Michael Brenndoerfer, Najmeh Sadoughi, James Fone, Nico Axtmann, Mark Miller, and David Suendermann-Oeft

A medical scribe is a clinical professional who charts patient—physician encounters in real time, relieving physicians of most of their administrative burden and substantially increasing productivity and job satisfaction. We present a complete implementation of an automated medical scribe. Our system can serve either as a scalable, standardized, and economical alternative to human scribes; or as an assistive tool for them, providing

a first draft of a report along with a convenient means to modify it. This solution is, to our knowledge, the first automated scribe ever presented and relies upon multiple speech and language technologies, including speaker diarization, medical speech recognition, knowledge extraction, and natural language generation.

CL Scholar: The ACL Anthology Knowledge Graph Miner

Mayank Singh, Pradeep Dogga, Sohan Patro, Dhiraj Barnwal, Ritam Dutt, Rajarshi Haldar, Pawan Goyal, and Animesh Mukherjee

We present CL Scholar, the ACL Anthology knowledge graph miner to facilitate high-quality search and exploration of current research progress in the computational linguistics community. In contrast to previous works, periodically crawling, indexing and processing of new incoming articles is completely automated in the current system. CL Scholar utilizes both textual and network information for knowledge graph construction. As an additional novel initiative, CL Scholar supports more than 1200 scholarly natural language queries along with standard keyword-based search on constructed knowledge graph. It answers binary, statistical and list based natural language queries. The current system is deployed at <http://cnrg.iitkgp.ac.in/aclakg>. We also provide REST API support along with bulk download facility. Our code and data are available at <https://github.com/CLScholar>.

Recruitment Lunch

The Recruitment Lunch is meant for senior graduate students seeking to make employment connections with our conference sponsors and others. Join us for a free lunch on Saturday, June 2nd, at 12:30, in the Elite Hall A at the hotel. Entrance tickets will be included in your registration materials or, if not a student, you may sign up to attend this event at the registration desk (our limit is 400 people + Recruiters). Get to know others who share similar interests and goals and may become your lifelong colleagues.

Industry Keynote Address: Mari Ostendorf

Building a Socialbot: Lessons Learned from 10M Conversations

Saturday, June 2, 2018, 14:00–15:00

Chair: Yunyao Li

Empire Ballroom

Abstract: Researchers in artificial intelligence have long been interested in the challenge of developing a system that can converse with humans, but much of the research has focused on text-based interactions and constrained contexts. This talk looks at open domain, spoken interactions enabled by a socialbot – Sounding Board – designed to be a conversational gateway to the web. The recent Alexa Prize competition made it possible for the Sounding Board team to develop a socialbot, learning from millions of conversations with real users. This talk will describe the system architecture and what we learned from all these interactions, emphasizing issues related to working with speech, online content and user variation that impact future directions for the field and for university-industry partnerships.

Biography: Mari is a professor of Electrical Engineering and Associate Vice Provost for Research at the University of Washington. She is a Fellow of IEEE and ISCA, and winner of the 2018 IEEE James L. Flanagan Speech and Audio Processing Award. Her recent research has emphasized social communication, and she served as a faculty advisor for the student team winning the inaugural AlexaPrize competition to build a socialbot.

Session 3 Overview – Saturday, June 2, 2018

Track A	Track B <i>Machine Learning 2</i>	Track C <i>Social Media and Computational Social Science 1</i>	Track D <i>Vision, Robotics and Other Grounding 1</i>	Track P <i>Industry: Panel</i>
Empire A	Empire B	Empire C	Empire D	Elite Hall B
Posters and Demos Deep Dirichlet Multinomial Regression <i>A. Benton and M. Dredze</i>	Microblog Conversation Recommendation via Joint Modeling of Topics and Discourse <i>X. Zeng, J. Li, L. Wang, N. Beauchamp, S. Shugars, and K.-F. Wong</i>	Scene Graph Parsing as Dependency Parsing <i>Y.-S. Wang, C. Liu, X. Zeng, and A. Yuille</i>		
Training Structured Prediction Energy Networks with Indirect Supervision <i>A. Rooshenas, A. Kamath, and A. McCallum</i>	Before Name-Calling: Dynamics and Triggers of Ad Hominem Fallacies in Web Argumentation <i>I. Habernal, H. Wachsmuth, I. Gurevych, and B. Stein</i>	Learning Visually Grounded Sentence Representations <i>D. Kiela, A. Conneau, A. Jabri, and M. Nickel</i>	Industry Panel: Careers in the Industry	Semantics 1 / Sentiment Analysis 1 / Demos
Anchored Correlation Explanation: Topic Modeling with Minimal Domain Knowledge (TACL) <i>R. Gallagher, K. Reing, D. Kale, and G. Ver Steeg</i>	Si O No, Que Pensé? Catalonian Independence and Linguistic Identity on Social Media <i>I. Stewart, Y. Pinter, and J. Eisenstein</i>	Comparatives, Quantifiers, Proportions: a Multi-Task Model for the Learning of Quantities from Vision <i>S. Pezzelle, I.-T. Sorodoc, and R. Bernardi</i>		16:06
Aspect-augmented Adversarial Networks for Domain Adaptation (TACL) <i>Y. Zhang, R. Barzilay, and T. Jaakkola</i>	Multiple Instance Learning Networks for Fine-Grained Sentiment Analysis (TACL) <i>S. Angelidis and M. Lapata</i>	Being Negative but Constructively: Lessons Learnt from Creating Better Visual Question Answering Datasets <i>W.-L. Chao, H. Hu, and F. Sha</i>		16:24

Parallel Session 3

Oral: Machine Learning 2

Empire A

Chair: Hal Daumé III

Deep Dirichlet Multinomial Regression

Adrian Benton and Mark Dredze

15:30–15:47

Dirichlet Multinomial Regression (DMR) and other supervised topic models can incorporate arbitrary document-level features to inform topic priors. However, their ability to model corpora are limited by the representation and selection of these features – a choice the topic modeler must make. Instead, we seek models that can learn the feature representations upon which to condition topic selection. We present deep Dirichlet Multinomial Regression (dDMR), a generative topic model that simultaneously learns document feature representations and topics. We evaluate dDMR on three datasets: New York Times articles with fine-grained tags, Amazon product reviews with product images, and Reddit posts with subreddit identity. dDMR learns representations that outperform DMR and LDA according to heldout perplexity and are more effective at downstream predictive tasks as the number of topics grows. Additionally, human subjects judge dDMR topics as being more representative of associated document features. Finally, we find that supervision leads to faster convergence as compared to an LDA baseline and that dDMR's model fit is less sensitive to training parameters than DMR.

Training Structured Prediction Energy Networks with Indirect Supervision

Amirmohammad Rooshenas, Aishwarya Kamath, and Andrew McCallum

15:48–16:05

This paper introduces rank-based training of structured prediction energy networks (SPENs). Our method samples from output structures using gradient descent and minimizes the ranking violation of the sampled structures with respect to a scalar scoring function defined with domain knowledge. We have successfully trained SPEN for citation field extraction without any labeled data instances, where the only source of supervision is a simple human-written scoring function. Such scoring functions are often easy to provide; the SPEN then furnishes an efficient structured prediction inference procedure.

Anchored Correlation Explanation: Topic Modeling with Minimal Domain Knowledge (TACL)

Ryan Gallagher, Kyle Reing, David Kale, and Greg Ver Steeg

16:06–16:23

While generative models such as Latent Dirichlet Allocation (LDA) have proven fruitful in topic modeling, they often require detailed assumptions and careful specification of hyperparameters. Such model complexity issues only compound when trying to generalize generative models to incorporate human input. We introduce Correlation Explanation (CorEx), an alternative approach to topic modeling that does not assume an underlying generative model, and instead learns maximally informative topics through an information-theoretic framework. This framework naturally generalizes to hierarchical and semi-supervised extensions with no additional modeling assumptions. In particular, word-level domain knowledge can be flexibly incorporated within CorEx through anchor words, allowing topic separability and representation to be promoted with minimal human intervention. Across a variety of datasets, metrics, and experiments, we demonstrate that CorEx produces topics that are comparable in quality to those produced by unsupervised and semi-supervised variants of LDA.

Aspect-augmented Adversarial Networks for Domain Adaptation (TACL)

Yuan Zhang, Regina Barzilay, and Tommi Jaakkola

16:24–16:42

We introduce a neural method for transfer learning between two (source and target) classification tasks or aspects over the same domain. Rather than training on target labels, we use a few keywords pertaining to source and target aspects indicating sentence relevance instead of document class labels. Documents are encoded by learning to embed and softly select relevant sentences in an aspect-dependent manner. A shared classifier is trained on the source encoded documents and labels, and applied to target encoded documents. We ensure transfer through aspect-adversarial training so that encoded documents are, as sets, aspect-invariant. Experimental results demonstrate that our approach outperforms different baselines and model variants on two datasets, yielding an improvement of 27% on a pathology dataset and 5% on a review dataset.

Oral: Social Media and Computational Social Science 1

Empire B

*Chair: Margaret Mitchell***Microblog Conversation Recommendation via Joint Modeling of Topics and Discourse***Xingshan Zeng, Jing Li, Lu Wang, Nicholas Beauchamp, Sarah Shugars, and Kam-Fai Wong*
15:30–15:47

Millions of conversations are generated every day on social media platforms. With limited attention, it is challenging for users to select which discussions they would like to participate in. Here we propose a new method for microblog conversation recommendation. While much prior work has focused on post-level recommendation, we exploit both the conversational context, and user content and behavior preferences. We propose a statistical model that jointly captures: (1) topics for representing user interests and conversation content, and (2) discourse modes for describing user replying behavior and conversation dynamics. Experimental results on two Twitter datasets demonstrate that our system outperforms methods that only model content without considering discourse.

Before Name-Calling: Dynamics and Triggers of Ad Hominem Fallacies in Web Argumentation*Ivan Habernal, Henning Wachsmuth, Iryna Gurevych, and Benno Stein*

15:48–16:05

Arguing without committing a fallacy is one of the main requirements of an ideal debate. But even when debating rules are strictly enforced and fallacious arguments punished, arguers often lapse into attacking the opponent by an ad hominem argument. As existing research lacks solid empirical investigation of the typology of ad hominem arguments as well as their potential causes, this paper fills this gap by (1) performing several large-scale annotation studies, (2) experimenting with various neural architectures and validating our working hypotheses, such as controversy or reasonableness, and (3) providing linguistic insights into triggers of ad hominem using explainable neural network architectures.

Si O No, Que Penses? Catalonian Independence and Linguistic Identity on Social Media*Ian Stewart, Yuval Pinter, and Jacob Eisenstein*

16:06–16:23

Political identity is often manifested in language variation, but the relationship between the two is still relatively unexplored from a quantitative perspective. This study examines the use of Catalan, a language local to the semi-autonomous region of Catalonia in Spain, on Twitter in discourse related to the 2017 independence referendum. We corroborate prior findings that pro-independence tweets are more likely to include the local language than anti-independence tweets. We also find that Catalan is used more often in referendum-related discourse than in other contexts, contrary to prior findings on language variation. This suggests a strong role for the Catalan language in the expression of Catalonian political identity.

Multiple Instance Learning Networks for Fine-Grained Sentiment Analysis (TACL)*Stefanos Angelidis and Mirella Lapata*

16:24–16:42

We consider the task of fine-grained sentiment analysis from the perspective of multiple instance learning (MIL). Our neural model is trained on document sentiment labels, and learns to predict the sentiment of text segments, i.e. sentences or elementary discourse units (EDUs), without segment-level supervision. We introduce an attention-based polarity scoring method for identifying positive and negative text snippets and a new dataset which we call SPOT (as shorthand for Segment-level POlariTy annotations) for evaluating MIL-style sentiment models like ours. Experimental results demonstrate superior performance against multiple baselines, whereas a judgement elicitation study shows that EDU-level opinion extraction produces more informative summaries than sentence-based alternatives.

Oral: Vision, Robotics and Other Grounding 1

Empire C

*Chair: Mirella Lapata***Scene Graph Parsing as Dependency Parsing***Yu-Siang Wang, Chenxi Liu, Xiaohui Zeng, and Alan Yuille*

15:30–15:47

In this paper, we study the problem of parsing structured knowledge graphs from textual descriptions. In particular, we consider the scene graph representation that considers objects together with their attributes and relations: this representation has been proved useful across a variety of vision and language applications. We begin by introducing an alternative but equivalent edge-centric view of scene graphs that connect to dependency parses. Together with a careful redesign of label and action space, we combine the two-stage pipeline

used in prior work (generic dependency parsing followed by simple post-processing) into one, enabling end-to-end training. The scene graphs generated by our learned neural dependency parser achieve an F-score similarity of 49.67% to ground truth graphs on our evaluation set, surpassing best previous approaches by 5%. We further demonstrate the effectiveness of our learned parser on image retrieval applications.

Learning Visually Grounded Sentence Representations

Douwe Kiela, Alexis Conneau, Allan Jabri, and Maximilian Nickel

15:48–16:05

We investigate grounded sentence representations, where we train a sentence encoder to predict the image features of a given caption—i.e., we try to “imagine” how a sentence would be depicted visually—and use the resultant features as sentence representations. We examine the quality of the learned representations on a variety of standard sentence representation quality benchmarks, showing improved performance for grounded models over non-grounded ones. In addition, we thoroughly analyze the extent to which grounding contributes to improved performance, and show that the system also learns improved word embeddings.

Comparatives, Quantifiers, Proportions: a Multi-Task Model for the Learning of Quantities from Vision

Sandro Pezzelle, Ionut-Teodor Sorodoc, and Raffaella Bernardi

16:06–16:23

The present work investigates whether different quantification mechanisms (set comparison, vague quantification, and proportional estimation) can be jointly learned from visual scenes by a multi-task computational model. The motivation is that, in humans, these processes underlie the same cognitive, non-symbolic ability, which allows an automatic estimation and comparison of set magnitudes. We show that when information about lower-complexity tasks is available, the higher-level proportional task becomes more accurate than when performed in isolation. Moreover, the multi-task model is able to generalize to unseen combinations of target/non-target objects. Consistently with behavioral evidence showing the interference of absolute number in the proportional task, the multi-task model no longer works when asked to provide the number of target objects in the scene.

Being Negative but Constructively: Lessons Learnt from Creating Better Visual Question Answering Datasets

Wei-Lun Chao, Hexiang Hu, and Fei Sha

16:24–16:42

Visual question answering (Visual QA) has attracted a lot of attention lately, seen essentially as a form of (visual) Turing test that artificial intelligence should strive to achieve. In this paper, we study a crucial component of this task: how can we design good datasets for the task? We focus on the design of multiple-choice based datasets where the learner has to select the right answer from a set of candidate ones including the target (i.e., the correct one) and the decoys (i.e., the incorrect ones). Through careful analysis of the results attained by state-of-the-art learning models and human annotators on existing datasets, we show that the design of the decoy answers has a significant impact on how and what the learning models learn from the datasets. In particular, the resulting learner can ignore the visual information, the question, or both while still doing well on the task. Inspired by this, we propose automatic procedures to remedy such design deficiencies. We apply the procedures to re-construct decoy answers for two popular Visual QA datasets as well as to create a new Visual QA dataset from the Visual Genome project, resulting in the largest dataset for this task. Extensive empirical studies show that the design deficiencies have been alleviated in the remedied datasets and the performance on them is likely a more faithful indicator of the difference among learning models. The datasets are released and publicly available via http://www.teds.usc.edu/website_vqa/.

Industry Panel: Careers in the Industry

Empire D

Chair: Philip Resnik

Posters and Demos: Session 3

Posters: Semantics 1

Time: 15:30–17:00

Chair: Scott Wen-tau Yih

Location: Elite Hall B

A Transition-Based Algorithm for Unrestricted AMR Parsing

David Vilares and Carlos Gómez-Rodríguez

Non-projective parsing can be useful to handle cycles and reentrancy in AMR graphs. We explore this idea and introduce a greedy left-to-right non-projective transition-based parser. At each parsing configuration, an oracle decides whether to create a concept or whether to connect a pair of existing concepts. The algorithm handles reentrancy and arbitrary cycles natively, i.e. within the transition system itself. The model is evaluated on the LDC2015E86 corpus, obtaining results close to the state of the art, including a Smatch of 64%, and showing good behavior on reentrant edges.

Abstract Meaning Representation for Paraphrase Detection

Fuad Issa, Marco Damonte, Shay B. Cohen, Xiaohui Yan, and Yi Chang

Abstract Meaning Representation (AMR) parsing aims at abstracting away from the syntactic realization of a sentence, and denote only its meaning in a canonical form. As such, it is ideal for paraphrase detection, a problem in which one is required to specify whether two sentences have the same meaning. We show that naïve use of AMR in paraphrase detection is not necessarily useful, and turn to describe a technique based on latent semantic analysis in combination with AMR parsing that significantly advances state-of-the-art results in paraphrase detection for the Microsoft Research Paraphrase Corpus. Our best results in the transductive setting are 86.6% for accuracy and 90.0% for F_1 measure.

Analogies in Complex Verb Meaning Shifts: the Effect of Affect in Semantic Similarity Models

Maximilian Köper and Sabine Schulte im Walde

We present a computational model to detect and distinguish analogies in meaning shifts between German base and complex verbs. In contrast to corpus-based studies, a novel dataset demonstrates that “regular” shifts represent the smallest class. Classification experiments relying on a standard similarity model successfully distinguish between four types of shifts, with verb classes boosting the performance, and affective features for abstractness, emotion and sentiment representing the most salient indicators.

attr2vec: Jointly Learning Word and Contextual Attribute Embeddings with Factorization Machines

Fabio Petroni, Vassilis Plachouras, Timothy Nugent, and Jochen L. Leidner

The widespread use of word embeddings is associated with the recent successes of many natural language processing (NLP) systems. The key approach of popular models such as word2vec and GloVe is to learn dense vector representations from the context of words. More recently, other approaches have been proposed that incorporate different types of contextual information, including topics, dependency relations, n-grams, and sentiment. However, these models typically integrate only limited additional contextual information, and often in ad hoc ways. In this work, we introduce attr2vec, a novel framework for jointly learning embeddings for words and contextual attributes based on factorization machines. We perform experiments with different types of contextual information. Our experimental results on a text classification task demonstrate that using attr2vec to jointly learn embeddings for words and Part-of-Speech (POS) tags improves results compared to learning the embeddings independently. Moreover, we use attr2vec to train dependency-based embeddings and we show that they exhibit higher similarity between functionally related words compared to traditional approaches.

Can Network Embedding of Distributional Thesaurus Be Combined with Word Vectors for Better Representation?

Abhik Jana and Pawan Goyal

Distributed representations of words learned from text have proved to be successful in various natural language processing tasks in recent times. While some methods represent words as vectors computed from text using predictive model (Word2vec) or dense count based model (GloVe), others attempt to represent these in a distributional thesaurus network structure where the neighborhood of a word is a set of words having adequate context overlap. Being motivated by recent surge of research in network embedding techniques (DeepWalk,

LINE, node2vec etc.), we turn a distributional thesaurus network into dense word vectors and investigate the usefulness of distributional thesaurus embedding in improving overall word representation. This is the first attempt where we show that combining the proposed word representation obtained by distributional thesaurus embedding with the state-of-the-art word representations helps in improving the performance by a significant margin when evaluated against NLP tasks like word similarity and relatedness, synonym detection, analogy detection. Additionally, we show that even without using any handcrafted lexical resources we can come up with representations having comparable performance in the word similarity and relatedness tasks compared to the representations where a lexical resource has been used.

Character-Based Neural Networks for Sentence Pair Modeling

Wuwei Lan and Wei Xu

Sentence pair modeling is critical for many NLP tasks, such as paraphrase identification, semantic textual similarity, and natural language inference. Most state-of-the-art neural models for these tasks rely on pretrained word embedding and compose sentence-level semantics in varied ways; however, few works have attempted to verify whether we really need pretrained embeddings in these tasks. In this paper, we study how effective subword-level (character and character n-gram) representations are in sentence pair modeling. Though it is well-known that subword models are effective in tasks with single sentence input, including language modeling and machine translation, they have not been systematically studied in sentence pair modeling tasks where the semantic and string similarities between texts matter. Our experiments show that subword models without any pretrained word embedding can achieve new state-of-the-art results on two social media datasets and competitive results on news data for paraphrase identification.

Deep Neural Models of Semantic Shift

Alex Rosenfeld and Katrin Erk

Diachronic distributional models track changes in word use over time. In this paper, we propose a deep neural network diachronic distributional model. Instead of modeling lexical change via a time series as is done in previous work, we represent time as a continuous variable and model a word's usage as a function of time. Additionally, we have also created a novel synthetic task which measures a model's ability to capture the semantic trajectory. This evaluation quantitatively measures how well a model captures the semantic trajectory of a word over time. Finally, we explore how well the derivatives of our model can be used to measure the speed of lexical change.

Determining Event Durations: Models and Error Analysis

Alakananda Vempala, Eduardo Blanco, and Alexis Palmer

This paper presents models to predict event durations. We introduce aspectual features that capture deeper linguistic information than previous work, and experiment with neural networks. Our analysis shows that tense, aspect and temporal structure of the clause provide useful clues, and that an LSTM ensemble captures relevant context around the event.

Diachronic Usage Relatedness (DURel): A Framework for the Annotation of Lexical Semantic Change

Dominik Schlechtweg, Sabine Schulte im Walde, and Stefanie Eckmann

We propose a framework that extends synchronic polysemy annotation to diachronic changes in lexical meaning, to counteract the lack of resources for evaluating computational models of lexical semantic change. Our framework exploits an intuitive notion of semantic relatedness, and distinguishes between innovative and reductive meaning changes with high inter-annotator agreement. The resulting test set for German comprises ratings from five annotators for the relatedness of 1,320 use pairs across 22 target words.

Directional Skip-Gram: Explicitly Distinguishing Left and Right Context for Word Embeddings

Yan Song, Shuming Shi, Jing Li, and Haisong Zhang

In this paper, we present directional skip-gram (DSG), a simple but effective enhancement of the skip-gram model by explicitly distinguishing left and right context in word prediction. In doing so, a direction vector is introduced for each word, whose embedding is thus learned by not only word co-occurrence patterns in its context, but also the directions of its contextual words. Theoretical and empirical studies on complexity illustrate that our model can be trained as efficient as the original skip-gram model, when compared to other extensions of the skip-gram model. Experimental results show that our model outperforms others on different datasets in semantic (word similarity measurement) and syntactic (part-of-speech tagging) evaluations, respectively.

Discriminating between Lexico-Semantic Relations with the Specialization Tensor Model

Goran Glavaš and Ivan Vučić

We present a simple and effective feed-forward neural architecture for discriminating between lexico-semantic relations (synonymy, antonymy, hypernymy, and meronymy). Our Specialization Tensor Model (STM) simultaneously produces multiple different specializations of input distributional word vectors, tailored for predicting lexico-semantic relations for word pairs. STM outperforms more complex state-of-the-art architectures on two benchmark datasets and exhibits stable performance across languages. We also show that, if coupled with a bilingual distributional space, the proposed model can transfer the prediction of lexico-semantic relations to a resource-lean target language without any training data.

Distributional Inclusion Vector Embedding for Unsupervised Hypernymy Detection

Haw-Shiu Chang, Ziyun Wang, Luke Vilnis, and Andrew McCallum

Modeling hypernymy, such as poodle is-a dog, is an important generalization aid to many NLP tasks, such as entailment, relation extraction, and question answering. Supervised learning from labeled hypernym sources, such as WordNet, limits the coverage of these models, which can be addressed by learning hypernyms from unlabeled text. Existing unsupervised methods either do not scale to large vocabularies or yield unacceptably poor accuracy. This paper introduces distributional inclusion vector embedding (DIVE), a simple-to-implement unsupervised method of hypernym discovery via per-word non-negative vector embeddings which preserve the inclusion property of word contexts. In experimental evaluations more comprehensive than any previous literature of which we are aware—evaluating on 11 datasets using multiple existing as well as newly proposed scoring functions—we find that our method provides up to double the precision of previous unsupervised methods, and the highest average performance, using a much more compact word representation, and yielding many new state-of-the-art results.

Evaluating bilingual word embeddings on the long tail

Fabienne Braune, Viktor Hangya, Tobias Eder, and Alexander Fraser

Bilingual word embeddings are useful for bilingual lexicon induction, the task of mining translations of given words. Many studies have shown that bilingual word embeddings perform well for bilingual lexicon induction but they focused on frequent words in general domains. For many applications, bilingual lexicon induction of rare and domain-specific words is of critical importance. Therefore, we design a new task to evaluate bilingual word embeddings on rare words in different domains. We show that state-of-the-art approaches fail on this task and present simple new techniques to improve bilingual word embeddings for mining rare words. We release new gold standard datasets and code to stimulate research on this task.

Frustratingly Easy Meta-Embedding – Computing Meta-Embeddings by Averaging Source Word Embeddings

Joshua Coates and Danushka Bollegala

Creating accurate meta-embeddings from pre-trained source embeddings has received attention lately. Methods based on global and locally-linear transformation and concatenation have shown to produce accurate meta-embeddings. In this paper, we show that the arithmetic mean of two distinct word embedding sets yields a performant meta-embedding that is comparable or better than more complex meta-embedding learning methods. The result seems counter-intuitive given that vector spaces in different source embeddings are not comparable and cannot be simply averaged. We give insight into why averaging can still produce accurate meta-embedding despite the incomparability of the source vector spaces.

Introducing Two Vietnamese Datasets for Evaluating Semantic Models of (Dis-)Similarity and Relatedness

Kim Anh Nguyen, Sabine Schulte im Walde, and Ngoc Thang Vu

We present two novel datasets for the low-resource language Vietnamese to assess models of semantic similarity: ViCon comprises pairs of synonyms and antonyms across word classes, thus offering data to distinguish between similarity and dissimilarity. ViSim-400 provides degrees of similarity across five semantic relations, as rated by human judges. The two datasets are verified through standard co-occurrence and neural network models, showing results comparable to the respective English datasets.

Lexical Substitution for Evaluating Compositional Distributional Models

Maja Buljan, Sebastian Padó, and Jan Šnajder

Compositional Distributional Semantic Models (CDSMs) model the meaning of phrases and sentences in vector space. They have been predominantly evaluated on limited, artificial tasks such as semantic sentence similarity on hand-constructed datasets. This paper argues for lexical substitution (LexSub) as a means to evaluate CDSMs. LexSub is a more natural task, enables us to evaluate meaning composition at the level of

individual words, and provides a common ground to compare CDSMs with dedicated LexSub models. We create a LexSub dataset for CDSM evaluation from a corpus with manual “all-words” LexSub annotation. Our experiments indicate that the Practical Lexical Function CDSM outperforms simple component-wise CDSMs and performs on par with the context2vec LexSub model using the same context.

Mining Possessions: Existence, Type and Temporal Anchors

Dhivya Chinnappa and Eduardo Blanco

This paper presents a corpus and experiments to mine possession relations from text. Specifically, we target alienable and control possessions, and assign temporal anchors indicating when the possession holds between possessor and possessee. We present new annotations for this task, and experimental results using both traditional classifiers and neural networks. Results show that the three subtasks (predicting possession existence, possession type and temporal anchors) can be automated.

Mittens: an Extension of GloVe for Learning Domain-Specialized Representations

Nicholas Dingwall and Christopher Potts

We present a simple extension of the GloVe representation learning model that begins with general-purpose representations and updates them based on data from a specialized domain. We show that the resulting representations can lead to faster learning and better results on a variety of tasks.

Neural Tensor Networks with Diagonal Slice Matrices

Takahiro Ishihara, Katsuhiko Hayashi, Hitoshi Manabe, Masashi Shimbo, and Masaaki Nagata

Although neural tensor networks (NTNs) have been successful in many NLP tasks, they require a large number of parameters to be estimated, which often leads to overfitting and a long training time. We address these issues by applying eigendecomposition to each slice matrix of a tensor to reduce its number of parameters. First, we evaluate our proposed NTN models on knowledge graph completion. Second, we extend the models to recursive NTNs (RNTNs) and evaluate them on logical reasoning tasks. These experiments show that our proposed models learn better and faster than the original (R)NTNs.

Olive Oil is Made of Olives, Baby Oil is Made for Babies: Interpreting Noun Compounds Using Paraphrases in a Neural Model

Vered Shwartz and Chris Waterson

Automatic interpretation of the relation between the constituents of a noun compound, e.g. olive oil (source) and baby oil (purpose) is an important task for many NLP applications. Recent approaches are typically based on either noun-compound representations or paraphrases. While the former has initially shown promising results, recent work suggests that the success stems from memorizing single prototypical words for each relation. We explore a neural paraphrasing approach that demonstrates superior performance when such memorization is not possible.

Post-Specialisation: Retrofitting Vectors of Words Unseen in Lexical Resources

Ivan Vulić, Goran Glavaš, Nikola Mrkšić, and Anna Korhonen

Word vector specialisation (also known as retrofitting) is a portable, light-weight approach to fine-tuning arbitrary distributional word vector spaces by injecting external knowledge from rich lexical resources such as WordNet. By design, these post-processing methods only update the vectors of words occurring in external lexicons, leaving the representations of all unseen words intact. In this paper, we show that constraint-driven vector space specialisation can be extended to unseen words. We propose a novel post-specialisation method that: a) preserves the useful linguistic knowledge for seen words; while b) propagating this external signal to unseen words in order to improve their vector representations as well. Our post-specialisation approach explicits a non-linear specialisation function in the form of a deep neural network by learning to predict specialised vectors from their original distributional counterparts. The learned function is then used to specialise vectors of unseen words. This approach, applicable to any post-processing model, yields considerable gains over the initial specialisation models both in intrinsic word similarity tasks, and in two downstream tasks: dialogue state tracking and lexical text simplification. The positive effects persist across three languages, demonstrating the importance of specialising the full vocabulary of distributional word vector spaces.

Semantic Pleonasm Detection

Omid Kashefi, Andrew T. Lucas, and Rebecca Hwa

Pleonasms are words that are redundant. To aid the development of systems that detect pleonasms in text, we introduce an annotated corpus of semantic pleonasms. We validate the integrity of the corpus with interannotator agreement analyses. We also compare it against alternative resources in terms of their effects on several automatic redundancy detection methods.

Similarity Measures for the Detection of Clinical Conditions with Verbal Fluency Tasks*Felipe Paula, Rodrigo Wilkens, Marco Idiart, and Aline Villavicencio*

Semantic Verbal Fluency tests have been used in the detection of certain clinical conditions, like Dementia. In particular, given a sequence of semantically related words, a large number of switches from one semantic class to another has been linked to clinical conditions. In this work, we investigate three similarity measures for automatically identifying switches in semantic chains: semantic similarity from a manually constructed resource, and word association strength and semantic relatedness, both calculated from corpora. This information is used for building classifiers to distinguish healthy controls from clinical cases with early stages of Alzheimer's Disease and Mild Cognitive Deficits. The overall results indicate that for clinical conditions the classifiers that use these similarity measures outperform those that use a gold standard taxonomy.

Sluice Resolution without Hand-Crafted Features over Brittle Syntax Trees*Ola Rønning, Daniel Hardt, and Anders Søgaard*

Sluice resolution in English is the problem of finding antecedents of *wh*-fronted ellipses. Previous work has relied on hand-crafted features over syntax trees that scale poorly to other languages and domains; in particular, to dialogue, which is one of the most interesting applications of sluice resolution. Syntactic information is arguably important for sluice resolution, but we show that multi-task learning with partial parsing as auxiliary tasks effectively closes the gap and buys us an additional 9%-error reduction over previous work. Since we are not directly relying on features from partial parsers, our system is more robust to domain shifts, giving a 26%-error reduction on embedded sluices in dialogue.

The Word Analogy Testing Caveat*Natalie Schluter*

There are some important problems in the evaluation of word embeddings using standard word analogy tests. In particular, in virtue of the assumptions made by systems generating the embeddings, these remain tests over randomness. We show that even supposing there were such word analogy regularities that should be detected in the word embeddings obtained via unsupervised means, standard word analogy test implementation practices provide distorted or contrived results. We raise concerns regarding the use of Principal Component Analysis to 2 or 3 dimensions as a provision of visual evidence for the existence of word analogy relations in embeddings. Finally, we propose some solutions to these problems.

Transition-Based Chinese AMR Parsing*Chuan Wang, Bin Li, and Nianwen Xue*

This paper presents the first AMR parser built on the Chinese AMR bank. By applying a transition-based AMR parsing framework to Chinese, we first investigate how well the transitions first designed for English AMR parsing generalize to Chinese and provide a comparative analysis between the transitions for English and Chinese. We then perform a detailed error analysis to identify the major challenges in Chinese AMR parsing that we hope will inform future research in this area.

Unsupervised Learning of Sentence Embeddings Using Compositional n-Gram Features*Matteo Tagliardini, Prakhar Gupta, and Martin Jaggi*

The recent tremendous success of unsupervised word embeddings in a multitude of applications raises the obvious question if similar methods could be derived to improve embeddings (i.e. semantic representations) of word sequences as well. We present a simple but efficient unsupervised objective to train distributed representations of sentences. Our method outperforms the state-of-the-art unsupervised models on most benchmark tasks, highlighting the robustness of the produced general-purpose sentence embeddings.

Combining Abstractness and Language-specific Theoretical Indicators for Detecting Non-Literal Usage of Estonian Particle Verbs (SRW)*Eleri Aedmaa, Maximilian Köper, and Sabine Schulte im Walde*

This paper presents two novel datasets and a random-forest classifier to automatically predict literal vs. non-literal language usage for a highly frequent type of multi-word expression in a low-resource language, i.e., Estonian. We demonstrate the value of language-specific indicators induced from theoretical linguistic research, which outperform a high majority baseline when combined with language-independent features of non-literal language (such as abstractness).

Verb Alternations and Their Impact on Frame Induction (SRW)*Esther Seyffarth*

Frame induction is the automatic creation of frame-semantic resources similar to FrameNet or PropBank, which map lexical units of a language to frame representations of each lexical unit's semantics. For verbs, these representations usually include a specification of their argument slots and of the selectional restrictions

that apply to each slot. Verbs that participate in diathesis alternations have different syntactic realizations whose semantics are closely related, but not identical. We discuss the influence that such alternations have on frame induction, compare several possible frame structures for verbs in the causative alternation, and propose a systematic analysis of alternating verbs that encodes their similarities as well as their differences.

A Generalized Knowledge Hunting Framework for the Winograd Schema Challenge (SRW)

Ali Emami, Adam Trischler, Kaheer Suleman, and Jackie Chi Kit Cheung

We introduce an automatic system that performs well on two common-sense reasoning tasks, the Winograd Schema Challenge (WSC) and the Choice of Plausible Alternatives (COPA). Problem instances from these tasks require diverse, complex forms of inference and knowledge to solve. Our method uses a knowledge-hunting module to gather text from the web, which serves as evidence for candidate problem resolutions. Given an input problem, our system generates relevant queries to send to a search engine. It extracts and classifies knowledge from the returned results and weighs it to make a resolution. Our approach improves F1 performance on the WSC by 0.16 over the previous best and is competitive with the state-of-the-art on COPA, demonstrating its general applicability.

Towards Qualitative Word Embeddings Evaluation: Measuring Neighbors Variation (SRW)

Benedicte Pierrejean and Ludovic Tanguy

We propose a method to study the variation lying between different word embeddings models trained with different parameters. We explore the variation between models trained with only one varying parameter by observing the distributional neighbors variation and show how changing only one parameter can have a massive impact on a given semantic space. We show that the variation is not affecting all words of the semantic space equally. Variation is influenced by parameters such as setting a parameter to its minimum or maximum value but it also depends on the corpus intrinsic features such as the frequency of a word. We identify semantic classes of words remaining stable across the models trained and specific words having high variation.

A Deeper Look into Dependency-Based Word Embeddings (SRW)

Sean MacAvaney and Amir Zeldes

We investigate the effect of various dependency-based word embeddings on distinguishing between functional and domain similarity, word similarity rankings, and two downstream tasks in English. Variations include word embeddings trained using context windows from Stanford and Universal dependencies at several levels of enhancement (ranging from unlabeled, to Enhanced++ dependencies). Results are compared to basic linear contexts and evaluated on several datasets. We found that embeddings trained with Universal and Stanford dependency contexts excel at different tasks, and that enhanced dependencies often improve performance.

Unsupervised Word Mapping Using Structural Similarities in Monolingual Embeddings (TACL)

Hanan Aldarmaki, Mahesh Mohan, and Mona Diab

Most existing methods for automatic bilingual dictionary induction rely on prior alignments between the source and target languages, such as parallel corpora or seed dictionaries. For many language pairs, such supervised alignments are not readily available. We propose an unsupervised approach for learning a bilingual dictionary for a pair of languages given their independently-learned monolingual word embeddings. The proposed method exploits local and global structures in monolingual vector spaces to align them such that similar words are mapped to each other. We show empirically that the performance of bilingual correspondents learned using our proposed unsupervised method is comparable to that of using supervised bilingual correspondents from a seed dictionary.

Posters: Sentiment Analysis 1

Time: 15:30–17:00

Location: Elite Hall B

Chair: Scott Wen-tau Yih

Knowledge-Enriched Two-Layered Attention Network for Sentiment Analysis

Abhishek Kumar, Daisuke Kawahara, and Sadao Kurohashi

We propose a novel two-layered attention network based on Bidirectional Long Short-Term Memory for sentiment analysis. The novel two-layered attention network takes advantage of the external knowledge bases to improve the sentiment prediction. It uses the Knowledge Graph Embedding generated using the WordNet. We build our model by combining the two-layered attention network with the supervised model based on Support

Vector Regression using a Multilayer Perceptron network for sentiment analysis. We evaluate our model on the benchmark dataset of SemEval 2017 Task 5. Experimental results show that the proposed model surpasses the top system of SemEval 2017 Task 5. The model performs significantly better by improving the state-of-the-art system at SemEval 2017 Task 5 by 1.7 and 3.7 points for sub-tracks 1 and 2 respectively.

Learning Domain Representation for Multi-Domain Sentiment Classification

Qi Liu, Yue Zhang, and Jiangming Liu

Training data for sentiment analysis are abundant in multiple domains, yet scarce for other domains. It is useful to leveraging data available for all existing domains to enhance performance on different domains. We investigate this problem by learning domain-specific representations of input sentences using neural network. In particular, a descriptor vector is learned for representing each domain, which is used to map adversarially trained domain-general Bi-LSTM input representations into domain-specific representations. Based on this model, we further expand the input representation with exemplary domain knowledge, collected by attending over a memory network of domain training data. Results show that our model outperforms existing methods on multi-domain sentiment analysis significantly, giving the best accuracies on two different benchmarks.

Learning Sentence Representations over Tree Structures for Target-dependent Classification

Junwen Duan, Xiao Ding, and Ting Liu

Target-dependent classification tasks, such as aspect-level sentiment analysis, perform fine-grained classifications towards specific targets. Semantic compositions over tree structures are promising for such tasks, as they can potentially capture long-distance interactions between targets and their contexts. However, previous work that operates on tree structures resorts to syntactic parsers or Treebank annotations, which are either subject to noise in informal texts or highly expensive to obtain. To address above issues, we propose a reinforcement learning based approach, which automatically induces target-specific sentence representations over tree structures. The underlying model is a RNN encoder-decoder that explores possible binary tree structures and a reward mechanism that encourages structures that improve performances on downstream tasks. We evaluate our approach on two benchmark tasks: firm-specific cumulative abnormal return prediction (based on formal news texts) and aspect-level sentiment analysis (based on informal social media texts). Experimental results show that our model gives superior performances compared to previous work that operates on parsed trees. Moreover, our approach gives some intuitions on how target-specific sentence representations can be achieved from its word constituents.

Letting Emotions Flow: Success Prediction by Modeling the Flow of Emotions in Books

Suraj Maharjan, Sudipta Kar, Manuel Montes, Fabio A. González, and Thamar Solorio

Books have the power to make us feel happiness, sadness, pain, surprise, or sorrow. An author's dexterity in the use of these emotions captivates readers and makes it difficult for them to put the book down. In this paper, we model the flow of emotions over a book using recurrent neural networks and quantify its usefulness in predicting success in books. We obtained the best weighted F1-score of 69% for predicting books' success in a multitask setting (simultaneously predicting success and genre of books).

Modeling Inter-Aspect Dependencies for Aspect-Based Sentiment Analysis

Devamanyu Hazarika, Soujanya Poria, Prateek Vir, Gangeshwar Krishnamurthy, Erik Cambria, and Roger Zimmermann

Aspect-based Sentiment Analysis is a fine-grained task of sentiment classification for multiple aspects in a sentence. Present neural-based models exploit aspect and its contextual information in the sentence but largely ignore the inter-aspect dependencies. In this paper, we incorporate this pattern by simultaneous classification of all aspects in a sentence along with temporal dependency processing of their corresponding sentence representations using recurrent networks. Results on the benchmark SemEval 2014 dataset suggest the effectiveness of our proposed approach.

Multi-Task Learning Framework for Mining Crowd Intelligence towards Clinical Treatment

Shweta Yadav, Asif Ekbal, Sriparna Saha, Pushpak Bhattacharyya, and Amit Sheth

In recent past, social media has emerged as an active platform in the context of healthcare and medicine. In this paper, we present a study where medical user's opinions on health-related issues are analyzed to capture the medical sentiment at a blog level. The medical sentiments can be studied in various facets such as medical condition, treatment, and medication that characterize the overall health status of the user. Considering these facets, we treat analysis of this information as a multi-task classification problem. In this paper, we adopt a novel adversarial learning approach for our multi-task learning framework to learn the sentiment's strengths

expressed in a medical blog. Our evaluation shows promising results for our target tasks.

Recurrent Entity Networks with Delayed Memory Update for Targeted Aspect-Based Sentiment Analysis*Fei Liu, Trevor Cohn, and Timothy Baldwin*

While neural networks have been shown to achieve impressive results for sentence-level sentiment analysis, targeted aspect-based sentiment analysis (TABSA) — extraction of fine-grained opinion polarity w.r.t. a pre-defined set of aspects — remains a difficult task. Motivated by recent advances in memory-augmented models for machine reading, we propose a novel architecture, utilising external “memory chains” with a delayed memory update mechanism to track entities. On a TABSA task, the proposed model demonstrates substantial improvements over state-of-the-art approaches, including those using external knowledge bases.

Relevant Emotion Ranking from Text Constrained with Emotion Relationships*Deyu Zhou, Yang Yang, and Yulan He*

Text might contain or invoke multiple emotions with varying intensities. As such, emotion detection, to predict multiple emotions associated with a given text, can be cast into a multi-label classification problem. We would like to go one step further so that a ranked list of relevant emotions are generated where top ranked emotions are more intensely associated with text compared to lower ranked emotions, whereas the rankings of irrelevant emotions are not important. A novel framework of relevant emotion ranking is proposed to tackle the problem. In the framework, the objective loss function is designed elaborately so that both emotion prediction and rankings of only relevant emotions can be achieved. Moreover, we observe that some emotions co-occur more often while other emotions rarely co-exist. Such information is incorporated into the framework as constraints to improve the accuracy of emotion detection. Experimental results on two real-world corpora show that the proposed framework can effectively deal with emotion detection and performs remarkably better than the state-of-the-art emotion detection approaches and multi-label learning methods.

Solving Data Sparsity for Aspect Based Sentiment Analysis Using Cross-Linguality and Multi-Linguality*Md Shad Akhtar, Palaash Sawant, Sukanta Sen, Asif Ekbal, and Pushpak Bhattacharyya*

Efficient word representations play an important role in solving various problems related to Natural Language Processing (NLP), data mining, text mining etc. The issue of data sparsity poses a great challenge in creating efficient word representation model for solving the underlying problem. The problem is more intensified in resource-poor scenario due to the absence of sufficient amount of corpus. In this work we propose to minimize the effect of data sparsity by leveraging bilingual word embeddings learned through a parallel corpus. We train and evaluate Long Short Term Memory (LSTM) based architecture for aspect level sentiment classification. The neural network architecture is further assisted by the hand-crafted features for the prediction. We show the efficacy of the proposed model against state-of-the-art methods in two experimental setups i.e. multi-lingual and cross-lingual.

SRL4ORL: Improving Opinion Role Labeling Using Multi-Task Learning with Semantic Role Labeling*Ana Marasovic and Anette Frank*

For over a decade, machine learning has been used to extract opinion-holder-target structures from text to answer the question “Who expressed what kind of sentiment towards what?”. Recent neural approaches do not outperform the state-of-the-art feature-based models for Opinion Role Labeling (ORL). We suspect this is due to the scarcity of labeled training data and address this issue using different multi-task learning (MTL) techniques with a related task which has substantially more data, i.e. Semantic Role Labeling (SRL). We show that two MTL models improve significantly over the single-task model for labeling of both holders and targets, on the development and the test sets. We found that the vanilla MTL model, which makes predictions using only shared ORL and SRL features, performs the best. With deeper analysis we determine what works and what might be done to make further improvements for ORL.

Benchmarks and Models for Entity-Oriented Polarity Detection (INDUSTRY)*Lidia Pivovarova, Arto Klami, and Roman Yangarber*

We address the problem of determining entity-oriented polarity in business news. This can be viewed as classifying the polarity of the sentiment expressed toward a given mention of a company in a news article. We present a complete, end-to-end approach to the problem. We introduce a new dataset of over 17,000 manually labeled documents, which is substantially larger than any currently available resources. We propose a benchmark solution based on convolutional neural networks for classifying entity-oriented polarity. Although our dataset is much larger than those currently available, it is small on the scale of datasets commonly used

for training robust neural network models. To compensate for this, we use transfer learning—pre-train the model on a much larger dataset, annotated for a related but different classification task, in order to learn a good representation for business text, and then fine-tune it on the smaller polarity dataset.

Demos

Time: 15:30–17:00

Location: Elite Hall B

ArgumenText: Searching for Arguments in Heterogeneous Sources

Christian Stab, Johannes Daxenberger, Chris Stahlhut, Tristan Miller, Benjamin Schiller, Christopher Tauchmann, Steffen Eger, and Iryna Gurevych

Argument mining is a core technology for enabling argument search in large corpora. However, most current approaches fall short when applied to heterogeneous texts. In this paper, we present an argument retrieval system capable of retrieving sentential arguments for any given controversial topic. By analyzing the highest-ranked results extracted from Web sources, we found that our system covers 89% of arguments found in expert-curated lists of arguments from an online debate portal, and also identifies additional valid arguments.

ClaimRank: Detecting Check-Worthy Claims in Arabic and English

Israa Jaradat, Pepa Gencheva, Alberto Barrón-Cedeño, Lluís Márquez, and Preslav Nakov

We present ClaimRank, an online system for detecting check-worthy claims. While originally trained on political debates, the system can work for any kind of text, e.g., interviews or just regular news articles. Its aim is to facilitate manual fact-checking efforts by prioritizing the claims that fact-checkers should consider first. ClaimRank supports both Arabic and English, it is trained on actual annotations from nine reputable fact-checking organizations (PolitiFact, FactCheck, ABC, CNN, NPR, NYT, Chicago Tribune, The Guardian, and Washington Post), and thus it can mimic the claim selection strategies for each and any of them, as well as for the union of them all.

360° Stance Detection

Sebastian Ruder, John Glover, Afshin Mehrabani, and Parsa Ghaffari

The proliferation of fake news and filter bubbles makes it increasingly difficult to form an unbiased, balanced opinion towards a topic. To ameliorate this, we propose 360° Stance Detection, a tool that aggregates news with multiple perspectives on a topic. It presents them on a spectrum ranging from support to opposition, enabling the user to base their opinion on multiple pieces of diverse evidence.

DebugSL: An Interactive Tool for Debugging Sentiment Lexicons

Andrew Schneider, John Male, Saroja Bhogadhi, and Eduard Dragut

We introduce DebugSL, a visual (Web) debugging tool for sentiment lexicons (SLs). Its core component implements our algorithms for the automatic detection of polarity inconsistencies in SLs. An inconsistency is a set of words and/or word-senses whose polarity assignments cannot all be simultaneously satisfied. DebugSL finds inconsistencies of small sizes in SLs and has a rich user interface which helps users in the correction process. The project source code is available at <https://github.com/atschneid/DebugSL>. A screencast of DebugSL can be viewed at <https://cis.temple.edu/~edragut/DebugSL.webm>.

Session 4 Overview – Saturday, June 2, 2018

	Track A <i>NLP Applications 2</i> Empire A	Track B <i>Question Answering 1</i> Empire B	Track C <i>SRW Highlight Papers</i> Empire C	Track D <i>Industry: Machine Translation</i> Empire D
17:00	Approaching Neural Grammatical Error Correction as a Low-Resource Machine Translation Task <i>M. Junczys-Dowmunt, R. Grundkiewicz, S. Guha, and K. Heafield</i>	Self-Training for Jointly Learning to Ask and Answer Questions <i>M. Sachan and E. Xing</i>	Igbo Diacritic Restoration using Embedding Models (SRW) <i>I. Ezeani, M. Hepple, I. Onyenwe, and E. Chioma</i> Using Classifier Features to Determine Language Transfer on Morphemes (SRW) <i>A. Lavrentovich</i>	Can Neural Machine Translation be Improved with User Feedback? (INDUSTRY) <i>J. Kreutzer, S. Khadivi, E. Matusov, and S. Riezler</i>
17:18	Robust Cross-Lingual Hypernymy Detection Using Dependency Context <i>S. Upadhyay, Y. Vyas, M. Carpuat, and D. Roth</i>	The Web as a Knowledge-base for Answering Complex Questions <i>A. Talmor and J. Berant</i>	End-to-End Learning of Task-Oriented Dialogs (SRW) <i>B. Liu</i>	Accelerating NMT Batched Beam Decoding with LMBR Posteriors for Deployment (INDUSTRY) <i>G. Iglesias, W. Tam-bellini, A. de Gispert, E. Hasler, and B. Byrne</i>
17:36	Noising and Denoising Natural Language: Diverse Backtranslation for Grammar Correction <i>Z. Xie, G. Gentithal, S. Xie, A. Ng, and D. Jurafsky</i>	A Meaning-Based Statistical English Math Word Problem Solver <i>C.-C. Liang, Y.-S. Wong, Y.-C. Lin, and K.-Y. Su</i>	Towards Generating Personalized Hospitalization Summaries (SRW) <i>S. Acharya, B. Di Eugenio, A. Boyd, R. Cameron, K. Dunn Lopez, P. Martyn-Nemeth, C. Dickens, and A. Ardati</i>	Pieces of Eight: 8-bit Neural Machine Translation (INDUSTRY) <i>J. Quinn and M. Balles-teros</i>
17:54	Towards Evaluating Narrative Quality In Student Writing (TACL) <i>S. Somasundaran, M. Flor, M. Chodorow, H. Molloy, B. Gyawali, and L. McCullla</i>	Strong Baselines for Simple Question Answering over Knowledge Graphs with and without Neural Networks <i>S. Mohammed, P. Shi, and J. Lin</i>	A Generalized Knowledge Hunting Framework for the Winograd Schema Challenge (SRW) <i>A. Emami, A. Trischler, K. Suleman, and J. C. K. Cheung</i>	From Dictations to Clinical Reports Using Machine Translation (INDUSTRY) <i>G. Finley, W. Salloum, N. Sadoighi, E. Edwards, A. Robinson, N. Axtmann, M. Brendnerfer, M. Miller, and D. Suendermann-Oeft</i>
18:13	Near Human-Level Performance in Grammatical Error Correction with Hybrid Machine Translation <i>R. Grundkiewicz and M. Junczys-Dowmunt</i>	Questionable Answers in Question Answering Research: Reproducibility and Variability of Published Results (TACL) <i>M. Crane</i>	Alignment, Acceptance, and Rejection of Group Identities in Online Political Discourse (SRW) <i>H. Shin and G. Doyle</i>	

Parallel Session 4

Oral: NLP Applications 2

Empire A

Chair: Yang Liu

Approaching Neural Grammatical Error Correction as a Low-Resource Machine Translation Task

Marcin Junczys-Dowmunt, Roman Grundkiewicz, Shubha Guha, and Kenneth Heafield 17:00–17:17

Previously, neural methods in grammatical error correction (GEC) did not reach state-of-the-art results compared to phrase-based statistical machine translation (SMT) baselines. We demonstrate parallels between neural GEC and low-resource neural MT and successfully adapt several methods from low-resource MT to neural GEC. We further establish guidelines for trustable results in neural GEC and propose a set of model-independent methods for neural GEC that can be easily applied in most GEC settings. Proposed methods include adding source-side noise, domain-adaptation techniques, a GEC-specific training-objective, transfer learning with monolingual data, and ensembling of independently trained GEC models and language models. The combined effects of these methods result in better than state-of-the-art neural GEC models that outperform previously best neural GEC systems by more than 10% M^2 on the CoNLL-2014 benchmark and 5.9% on the JFLEG test set. Non-neural state-of-the-art systems are outperformed by more than 2% on the CoNLL-2014 benchmark and by 4% on JFLEG.

Robust Cross-Lingual Hypernymy Detection Using Dependency Context

Shyam Upadhyay, Yogarshi Vyas, Marine Carpuat, and Dan Roth

17:18–17:35

Cross-lingual Hypernymy Detection involves determining if a word in one language (“fruit”) is a hypernym of a word in another language (“pomme” i.e. apple in French). The ability to detect hypernymy cross-lingually can aid in solving cross-lingual versions of tasks such as textual entailment and event coreference. We propose BiSparse-Dep, a family of unsupervised approaches for cross-lingual hypernymy detection, which learns sparse, bilingual word embeddings based on dependency contexts. We show that BiSparse-Dep can significantly improve performance on this task, compared to approaches based only on lexical context. Our approach is also robust, showing promise for low-resource settings: our dependency-based embeddings can be learned using a parser trained on related languages, with negligible loss in performance. We also crowd-source a challenging dataset for this task on four languages – Russian, French, Arabic, and Chinese. Our embeddings and datasets are publicly available.

Noising and Denoising Natural Language: Diverse Backtranslation for Grammar Correction

Ziang Xie, Guillaume Genthial, Stanley Xie, Andrew Ng, and Dan Jurafsky

17:36–17:53

Translation-based methods for grammar correction that directly map noisy, ungrammatical text to their clean counterparts are able to correct a broad range of errors; however, such techniques are bottlenecked by the need for a large parallel corpus of noisy and clean sentence pairs. In this paper, we consider synthesizing parallel data by noising a clean monolingual corpus. While most previous approaches introduce perturbations using features computed from local context windows, we instead develop error generation processes using a neural sequence transduction model trained to translate clean examples to their noisy counterparts. Given a corpus of clean examples, we propose beam search noising procedures to synthesize additional noisy examples that human evaluators were nearly unable to discriminate from nonsynthesized examples. Surprisingly, when trained on additional data synthesized using our best-performing noising scheme, our model approaches the same performance as when trained on additional nonsynthesized data.

Towards Evaluating Narrative Quality In Student Writing (TACL)

Swapna Somasundaran, Michael Flor, Martin Chodorow, Hillary Molloy, Binod Gyawali, and Laura McCulla 17:54–18:12

This work lays the foundation for automated assessments of narrative quality in student writing. We first manually score essays for narrative-relevant traits and sub-trait, and measure inter-annotator agreement. We then explore linguistic features that are indicative of good narrative writing and use them to build an automated scoring system. Experiments show that our features are more effective in scoring specific aspects of narrative quality than a state-of-the-art feature set.

Near Human-Level Performance in Grammatical Error Correction with Hybrid Machine Translation*Roman Grundkiewicz and Marcin Junczys-Dowmunt*

18:13–18:30

We combine two of the most popular approaches to automated Grammatical Error Correction (GEC): GEC based on Statistical Machine Translation (SMT) and GEC based on Neural Machine Translation (NMT). The hybrid system achieves new state-of-the-art results on the CoNLL-2014 and JFLEG benchmarks. This GEC system preserves the accuracy of SMT output and, at the same time, generates more fluent sentences as it typical for NMT. Our analysis shows that the created systems are closer to reaching human-level performance than any other GEC system reported so far.

Oral: Question Answering 1

Empire B

*Chair: Avirup Sil***Self-Training for Jointly Learning to Ask and Answer Questions***Mrinmaya Sachan and Eric Xing*

17:00–17:17

Building curious machines that can answer as well as ask questions is an important challenge for AI. The two tasks of question answering and question generation are usually tackled separately in the NLP literature. At the same time, both require significant amounts of supervised data which is hard to obtain in many domains. To alleviate these issues, we propose a self-training method for jointly learning to ask as well as answer questions, leveraging unlabeled text along with labeled question answer pairs for learning. We evaluate our approach on four benchmark datasets: SQuAD, MS MARCO, WikiQA and TrecQA, and show significant improvements over a number of established baselines on both question answering and question generation tasks. We also achieved new state-of-the-art results on two competitive answer sentence selection tasks: WikiQA and TrecQA.

The Web as a Knowledge-base for Answering Complex Questions*Alon Talmor and Jonathan Berant*

17:18–17:35

Answering complex questions is a time-consuming activity for humans that requires reasoning and integration of information. Recent work on reading comprehension made headway in answering simple questions, but tackling complex questions is still an ongoing research challenge. Conversely, semantic parsers have been successful at handling compositionality, but only when the information resides in a target knowledge-base. In this paper, we present a novel framework for answering broad and complex questions, assuming answering simple questions is possible using a search engine and a reading comprehension model. We propose to decompose complex questions into a sequence of simple questions, and compute the final answer from the sequence of answers. To illustrate the viability of our approach, we create a new dataset of complex questions, ComplexWebQuestions, and present a model that decomposes questions and interacts with the web to compute an answer. We empirically demonstrate that question decomposition improves performance from 20.8 precision1 to 27.5 precision1 on this new dataset.

A Meaning-Based Statistical English Math Word Problem Solver*Chao-Chun Liang, Yu-Shiang Wong, Yi-Chung Lin, and Keh-Yih Su*

17:36–17:53

We introduce MeSys, a meaning-based approach, for solving English math word problems (MWPs) via understanding and reasoning in this paper. It first analyzes the text, transforms both body and question parts into their corresponding logic forms, and then performs inference on them. The associated context of each quantity is represented with proposed role-tags (e.g., nsubj, verb, etc.), which provides the flexibility for annotating an extracted math quantity with its associated context information (i.e., the physical meaning of this quantity). Statistical models are proposed to select the operator and operands. A noisy dataset is designed to assess if a solver solves MWP mainly via understanding or mechanical pattern matching. Experimental results show that our approach outperforms existing systems on both benchmark datasets and the noisy dataset, which demonstrates that the proposed approach understands the meaning of each quantity in the text more.

Strong Baselines for Simple Question Answering over Knowledge Graphs with and without Neural Networks*Salman Mohammed, Peng Shi, and Jimmy Lin*

17:54–18:12

We examine the problem of question answering over knowledge graphs, focusing on simple questions that can be answered by the lookup of a single fact. Adopting a straightforward decomposition of the problem into entity detection, entity linking, relation prediction, and evidence combination, we explore simple yet strong baselines. On the popular SimpleQuestions dataset, we find that basic LSTMs and GRUs plus a few heuristics

yield accuracies that approach the state of the art, and techniques that do not use neural networks also perform reasonably well. These results show that gains from sophisticated deep learning techniques proposed in the literature are quite modest and that some previous models exhibit unnecessary complexity.

Questionable Answers in Question Answering Research: Reproducibility and Variability of Published Results (TACL)

Matt Crane

18:13–18:30

Based on theoretical reasoning it has been suggested that the reliability of findings published in the scientific literature decreases with the popularity of a research field" (Pfeiffer and Hoffmann, 2009). As we know, deep learning is very popular and the ability to reproduce results is an important part of science. There is growing concern within the deep learning community about the reproducibility of results that are presented. In this paper we present a number of controllable, yet unreported, effects that can substantially change the effectiveness of a sample model, and thusly the reproducibility of those results. Through these environmental effects we show that the commonly held belief that distribution of source code is all that is needed for reproducibility is not enough. Source code without a reproducible environment does not mean anything at all. In addition the range of results produced from these effects can be larger than the majority of incremental improvement reported.

Oral: SRW Highlight Papers

Empire C

Chairs: Umashanthi Pavalanathan, Shereen Oraby and Kyeongmin Rim

Igbo Diacritic Restoration using Embedding Models (SRW)

Ignatius Ezeani, Mark Hepple, Ikechukwu Onyenwe, and Enemoh Chioma

17:00–17:14

Igbo is a low-resource language spoken by approximately 30 million people worldwide. It is the native language of the Igbo people of south-eastern Nigeria. In Igbo language, diacritics - orthographic and tonal - play a huge role in the distinguishing the meaning and pronunciation of words. Omitting diacritics in texts often leads to lexical ambiguity. Diacritic restoration is a pre-processing task that replaces missing diacritics on words from which they have been removed. In this work, we applied embedding models to the diacritic restoration task and compared their performances to those of n-gram models. Although word embedding models have been successfully applied to various NLP tasks, it has not been used, to our knowledge, for diacritic restoration. Two classes of word embeddings models were used: those projected from the English embedding space; and those trained with Igbo bible corpus ($\approx 1m$). Our best result, 82.49%, is an improvement on the baseline n-gram models.

Using Classifier Features to Determine Language Transfer on Morphemes (SRW)

Alexandra Lavrentovich

17:15–17:29

The aim of this thesis is to perform a Native Language Identification (NLI) task where we identify an English learner's native language background based only on the learner's English writing samples. We focus on the use of English grammatical morphemes across four proficiency levels. The outcome of the computational task is connected to a position in second language acquisition research that holds all learners acquire English grammatical morphemes in the same order, regardless of native language background. We use the NLI task as a tool to uncover cross-linguistic influence on the developmental trajectory of morphemes. We perform a cross-corpus evaluation across proficiency levels to increase the reliability and validity of the linguistic features that predict the native language background. We include native English data to determine the different morpheme patterns used by native versus non-native English speakers. Furthermore, we conduct a human NLI task to determine the type and magnitude of language transfer cues used by human raters versus the classifier.

End-to-End Learning of Task-Oriented Dialogs (SRW)

Bing Liu

17:30–17:44

In this thesis proposal, we address the limitations of conventional pipeline design of task-oriented dialog systems and propose end-to-end learning solutions. We design neural network based dialog system that is able to robustly track dialog state, interface with knowledge bases, and incorporate structured query results into system responses to successfully complete task-oriented dialog. In learning such neural network based dialog systems, we propose hybrid offline training and online interactive learning methods. We introduce a multi-task learning method in pre-training the dialog agent in a supervised manner using task-oriented dialog corpora. The supervised training agent can further be improved via interacting with users and learning online from user demonstration and feedback with imitation and reinforcement learning. In addressing the sample efficiency issue with online policy learning, we further propose a method by combining the learning-from-user

and learning-from-simulation approaches to improve the online interactive learning efficiency.

Towards Generating Personalized Hospitalization Summaries (SRW)

Sabita Acharya, Barbara Di Eugenio, Andrew Boyd, Richard Cameron, Karen Dunn Lopez, Pamela Martyn-Nemeth, Carolyn Dickens, and Amer Ardati 17:45–17:59

Most of the health documents, including patient education materials and discharge notes, are usually flooded with medical jargons and contain a lot of generic information about the health issue. In addition, patients are only provided with the doctor's perspective of what happened to them in the hospital while the care procedure performed by nurses during their entire hospital stay is nowhere included. The main focus of this research is to generate personalized hospital-stay summaries for patients by combining information from physician discharge notes and nursing plan of care. It uses a metric to identify medical concepts that are Complex, extracts definitions for the concept from three external knowledge sources, and provides the simplest definition to the patient. It also takes various features of the patient into account, like their concerns and strengths, ability to understand basic health information, level of engagement in taking care of their health, and familiarity with the health issue and personalizes the content of the summaries accordingly. Our evaluation showed that the summaries contain 80% of the medical concepts that are considered as being important by both doctor and nurses. Three patient advisors (i.e. individuals who are trained in understanding patient experience extensively) verified the usability of our summaries and mentioned that they would like to get such summaries when they are discharged from hospital.

A Generalized Knowledge Hunting Framework for the Winograd Schema Challenge (SRW)

Ali Emami, Adam Trischler, Kaheer Suleman, and Jackie Chi Kit Cheung 18:00–18:14

We introduce an automatic system that performs well on two common-sense reasoning tasks, the Winograd Schema Challenge (WSC) and the Choice of Plausible Alternatives (COPA). Problem instances from these tasks require diverse, complex forms of inference and knowledge to solve. Our method uses a knowledge-hunting module to gather text from the web, which serves as evidence for candidate problem resolutions. Given an input problem, our system generates relevant queries to send to a search engine. It extracts and classifies knowledge from the returned results and weighs it to make a resolution. Our approach improves F1 performance on the WSC by 0.16 over the previous best and is competitive with the state-of-the-art on COPA, demonstrating its general applicability.

Alignment, Acceptance, and Rejection of Group Identities in Online Political Discourse (SRW)

Hagyeong Shin and Gabriel Doyle 18:15–18:30

Conversation is a joint social process, with participants cooperating to exchange information. This process is helped along through linguistic alignment: participants' adoption of each other's word use. This alignment is robust, appearing many settings, and is nearly always positive. We create an alignment model for examining alignment in Twitter conversations across antagonistic groups. This model finds that some word categories, specifically pronouns used to establish group identity and common ground, are negatively aligned. This negative alignment is observed despite other categories, which are less related to the group dynamics, showing the standard positive alignment. This suggests that alignment is strongly biased toward cooperative alignment, but that different linguistic features can show substantially different behaviors.

Industry: Machine Translation

Empire D

Chair: Kenneth Heafield

Can Neural Machine Translation be Improved with User Feedback? (INDUSTRY)

Julia Kreutzer, Shahram Khadivi, Evgeny Matusov, and Stefan Riezler 17:00–17:17

We present the first real-world application of methods for improving neural machine translation (NMT) with human reinforcement, based on explicit and implicit user feedback collected on the eBay e-commerce platform. Previous work has been confined to simulation experiments, whereas in this paper we work with real logged feedback for offline bandit learning of NMT parameters. We conduct a thorough analysis of the available explicit user judgments—five-star ratings of translation quality—and show that they are not reliable enough to yield significant improvements in bandit learning. In contrast, we successfully utilize implicit task-based feedback collected in a cross-lingual search task to improve task-specific and machine translation quality metrics.

Accelerating NMT Batched Beam Decoding with LMBR Posteriors for Deployment (INDUSTRY)*Gonzalo Iglesias, William Tambellini, Adrià de Gispert, Eva Hasler, and Bill Byrne* 17:18–17:35

We describe a batched beam decoding algorithm for NMT with LMBR n-gram posteriors, showing that LMBR techniques still yield gains on top of the best recently reported results with Transformers. We also discuss acceleration strategies for deployment, and the effect of the beam size and batching on memory and speed.

Pieces of Eight: 8-bit Neural Machine Translation (INDUSTRY)*Jerry Quinn and Miguel Ballesteros*

17:36–17:53

Neural machine translation has achieved levels of fluency and adequacy that would have been surprising a short time ago. Output quality is extremely relevant for industry purposes, however it is equally important to produce results in the shortest time possible, mainly for latency-sensitive applications and to control cloud hosting costs. In this paper we show the effectiveness of translating with 8-bit quantization for models that have been trained using 32-bit floating point values. Results show that 8-bit translation makes a non-negligible impact in terms of speed with no degradation in accuracy and adequacy.

From Dictations to Clinical Reports Using Machine Translation (INDUSTRY)*Gregory Finley, Wael Salloum, Najmeh Sadoughi, Erik Edwards, Amanda Robinson, Nico Axtmann, Michael Brenndoerfer, Mark Miller, and David Suendermann-Oeft* 17:54–18:12

A typical workflow to document clinical encounters entails dictating a summary, running speech recognition, and post-processing the resulting text into a formatted letter. Post-processing entails a host of transformations including punctuation restoration, truecasing, marking sections and headers, converting dates and numerical expressions, parsing lists, etc. In conventional implementations, most of these tasks are accomplished by individual modules. We introduce a novel holistic approach to post-processing that relies on machine cally-translation. We show how this technique outperforms an alternative conventional system—even learning to correct speech recognition errors during post-processing—while being much simpler to maintain.

4

Main Conference: Sunday, June 3

Overview

07:30 – 18:00	Registration				<i>Empire Foyer</i>			
07:45 – 08:45	Breakfast				<i>Empire Foyer</i>			
08:45 – 09:00	Announcements				<i>Empire Ballroom</i>			
09:00 – 10:00	Keynote Address: Kevin Knight (sponsored by Google)				<i>Empire Ballroom</i>			
10:00 – 10:30	Morning Coffee				<i>Empire Foyer</i>			
10:30 – 11:30	Research Oral 5	<i>Cognitive Modeling and Psycholinguistics 1</i>		<i>Summarization 1</i>	<i>Semantics 2</i>	Industry Oral 5	Poster/Demo 5/6 (10:30-12:00)	
		<i>Empire A</i>	<i>Empire B</i>	<i>Empire C</i>	<i>Empire D</i>	Machine Learning – Domain Adaptation		
11:30 – 12:30	Research Oral 6	<i>Machine Learning 3</i>	<i>Social Media and Computational Social Science 2</i>	<i>Vision, Robotics and Other Grounding 2</i>	<i>Named Entity and Language Generation</i>	Industry Oral 6	<i>Information Extraction 3 / Tagging, Chunking, Syntax and Parsing 1 / Demos</i>	
		<i>Empire A</i>	<i>Empire B</i>	<i>Empire C</i>	<i>Empire D</i>		<i>Elite Hall B</i>	
12:30 – 14:00	Lunch							
14:00 – 15:00	Industry Keynote Address: Daniel Marcu						<i>Empire Ballroom</i>	
15:00 – 15:30	Afternoon Coffee						<i>Empire Foyer</i>	
15:30 – 17:00	Research Oral 7	<i>Text Mining 1</i>	<i>Semantics 3</i>	<i>Tagging, Chunking, Syntax and Parsing 2</i>	<i>Ethics in NLP Panel</i>	Panel	Poster/Demo 7	
		<i>Empire A</i>	<i>Empire B</i>	<i>Empire C</i>	<i>Empire D</i>	Machine Learning 4 / Machine Translation 2 / Phonology, Morphology and Word Segmentation 2 / Demos		
17:00 – 17:15	Remembrance of Aravind Joshi						<i>Elite Hall B</i>	
17:15 – 18:30	Test Of Time						<i>Empire B</i>	
							<i>Empire B</i>	

Keynote Address: Kevin Knight (sponsored by Google)

The Moment When the Future Fell Asleep

Sunday, June 3, 2018, 9:00–10:00

Chair: Amanda Stent

Empire Ballroom

Biography: Kevin is a professor of computer science at the University of Southern California and fellow of the Information Sciences Institute. He is a 2014 fellow of the ACL for foundational contributions to machine translation, to the application of automata for NLP, to decipherment of historical manuscripts, to semantics and to generation.

Session 5 Overview – Sunday, June 3, 2018

Track A	Track B	Track C	Track D	Track P
<i>Cognitive Modeling and Psycholinguistics I</i> Empire A	<i>Summarization I</i> Empire B	<i>Semantics 2</i> Empire C	<i>Industry: Machine Learning – Domain Adaptation</i> Empire D	<i>Posters and Demos (10:30–12:00)</i> Elite Hall B
Fine-grained Temporal Orientation and its Relationship with Psychodemographic Correlates <i>S. Kamila, M. Hasanuzzaman, A. Ekbal, P. Bhattacharyya, and A. Way</i>	Semantic Structural Evaluation for Text Simplification <i>E. Sulem, O. Abend, and A. Rappoport</i>	Polyglot Semantic Parsing in APIs <i>K. Richardson, J. Berant, and J. Kuhn</i>	Selecting Machine-Translated Data for Quick Bootstrapping of a Natural Language Understanding System (INDUSTRY) <i>J. Gaspers, P. Karanasiou, and R. Chatterjee</i>	Information Extraction 3 / Tagging, Chunking, Syntax and Parsing 1 / Demos
Querying Word Embeddings for Similarity and Relatedness <i>F. Torabi Asr, R. Zinkov, and M. Jones</i>	Entity Commonsense Representation for Neural Abstractive Summarization <i>R. K. Amplayo, S. Lim, and S.-w. Hwang</i>	Modeling Semantic Plausibility by Injecting World Knowledge <i>S. Wang, G. Durrett, and K. Erk</i>	Fast and Scalable Expansion of Natural Language Understanding Functionality for Intelligent Agents (INDUSTRY) <i>A. K. Goyal, A. Metallinou, and S. Matsoukas</i>	10:48
Looking for Structure in Lexical and Acoustic-Prosodic Entrainment Behaviors <i>A. Weise and R. Levitan</i>	Newsroom: A Dataset of 1.3 Million Summaries with Diverse Extractive Strategies <i>M. Grusky, M. Naaman, and Y. Artzi</i>	Neural Models of Factuality <i>R. Rudinger, A. S. White, and B. Van Durme</i>	Bag of Experts Architectures for Model Reuse in Conversational Language Understanding (INDUSTRY) <i>R. Jha, A. Marin, S. Shivaprasad, and I. Zitouni</i>	11:06

Parallel Session 5

Oral: Cognitive Modeling and Psycholinguistics 1

Empire A

Chair: Kristy Hollingshead Seitz

Fine-grained Temporal Orientation and its Relationship with Psycho-demographic Correlates

Sabyasachi Kamila, Mohammed Hasanuzzaman, Asif Ekbal, Pushpak Bhattacharyya, and Andy Way 10:30–10:47

Temporal orientation refers to an individual's tendency to connect to the psychological concepts of past, present or future, and it affects personality, motivation, emotion, decision making and stress coping processes. The study of the social media users' psycho-demographic attributes from the perspective of human temporal orientation can be of utmost interest and importance to the business and administrative decision makers as it can provide an extra precious information for them to make informed decisions. In this paper, we propose a very first study to demonstrate the association between the sentiment view of the temporal orientation of the users and their different psycho-demographic attributes by analyzing their tweets. We first create a temporal orientation classifier in a minimally supervised way which classifies each tweet of the users in one of the three temporal categories, namely past, present, and future. A deep Bi-directional Long Short Term Memory (BLSTM) is used for the tweet classification task. Our tweet classifier achieves an accuracy of 78.27% when tested on a manually created test set. We then determine the users' overall temporal orientation based on their tweets on the social media. The sentiment is added to the tweets at the fine-grained level where each temporal tweet is given a sentiment with either of the positive, negative or neutral. Our experiment reveals that depending upon the sentiment view of temporal orientation, a user's attributes vary. We finally measure the correlation between the users' sentiment view of temporal orientation and their different psycho-demographic factors using regression.

Querying Word Embeddings for Similarity and Relatedness

Fatemeh Torabi Asr, Robert Zinkov, and Michael Jones 10:48–11:05

Word embeddings obtained from neural network models such as Word2Vec Skipgram have become popular representations of word meaning and have been evaluated on a variety of word similarity and relatedness norming data. Skipgram generates a set of word and context embeddings, the latter typically discarded after training. We demonstrate the usefulness of context embeddings in predicting asymmetric association between words from a recently published dataset of production norms (Jouravlev & McRae, 2016). Our findings suggest that humans respond with words closer to the cue within the context embedding space (rather than the word embedding space), when asked to generate thematically related words.

Looking for Structure in Lexical and Acoustic-Prosodic Entrainment Behaviors

Andreas Weise and Rivka Levitan 11:06–11:23

Entrainment has been shown to occur for various linguistic features individually. Motivated by cognitive theories regarding linguistic entrainment, we analyze speakers' overall entrainment behaviors and search for an underlying structure. We consider various measures of both acoustic-prosodic and lexical entrainment, measuring the latter with a novel application of two previously introduced methods in addition to a standard high-frequency word measure. We present a negative result of our search, finding no meaningful correlations, clusters, or principal components in various entrainment measures, and discuss practical and theoretical implications.

Oral: Summarization 1

Empire B

Chair: Rada Mihalcea

Semantic Structural Evaluation for Text Simplification

Elior Sulem, Omri Abend, and Ari Rappoport 10:30–10:47

Current measures for evaluating text simplification systems focus on evaluating lexical text aspects, neglecting its structural aspects. In this paper we propose the first measure to address structural aspects of text simplification, called SAMSA. It leverages recent advances in semantic parsing to assess simplification quality by decomposing the input based on its semantic structure and comparing it to the output. SAMSA provides a

reference-less automatic evaluation procedure, avoiding the problems that reference-based methods face due to the vast space of valid simplifications for a given sentence. Our human evaluation experiments show both SAMSA's substantial correlation with human judgments, as well as the deficiency of existing reference-based measures in evaluating structural simplification.

Entity Commonsense Representation for Neural Abstractive Summarization

Reinald Kim Amplayo, Seonjae Lim, and Seung-won Hwang

10:48–11:05

A major proportion of a text summary includes important entities found in the original text. These entities build up the topic of the summary. Moreover, they hold commonsense information once they are linked to a knowledge base. Based on these observations, this paper investigates the usage of linked entities to guide the decoder of a neural text summarizer to generate concise and better summaries. To this end, we leverage on an off-the-shelf entity linking system (ELS) to extract linked entities and propose Entity2Topic (E2T), a module easily attachable to a sequence-to-sequence model that transforms a list of entities into a vector representation of the topic of the summary. Current available ELS's are still not sufficiently effective, possibly introducing unresolved ambiguities and irrelevant entities. We resolve the imperfections of the ELS by (a) encoding entities with selective disambiguation, and (b) pooling entity vectors using firm attention. By applying E2T to a simple sequence-to-sequence model with attention mechanism as base model, we see significant improvements of the performance in the Gigaword (sentence to title) and CNN (long document to multi-sentence highlights) summarization datasets by at least 2 ROUGE points.

Newsroom: A Dataset of 1.3 Million Summaries with Diverse Extractive Strategies

Max Grusky, Mor Naaman, and Yoav Artzi

11:06–11:23

We present NEWSROOM, a summarization dataset of 1.3 million articles and summaries written by authors and editors in newsrooms of 38 major news publications. Extracted from search and social media metadata between 1998 and 2017, these high-quality summaries demonstrate high diversity of summarization styles. In particular, the summaries combine abstractive and extractive strategies, borrowing words and phrases from articles at varying rates. We analyze the extraction strategies used in NEWSROOM summaries against other datasets to quantify the diversity and difficulty of our new data, and train existing methods on the data to evaluate its utility and challenges. The dataset is available online at summar.es.

Oral: Semantics 2

Empire C

Chair: Jonathan May

Polyglot Semantic Parsing in APIs

Kyle Richardson, Jonathan Berant, and Jonas Kuhn

10:30–10:47

Traditional approaches to semantic parsing (SP) work by training individual models for each available parallel dataset of text-meaning pairs. In this paper, we explore the idea of polyglot semantic translation, or learning semantic parsing models that are trained on multiple datasets and natural languages. In particular, we focus on translating text to code signature representations using the software component datasets of Richardson and Kuhn (2017b,a). The advantage of such models is that they can be used for parsing a wide variety of input natural languages and output programming languages, or mixed input languages, using a single unified model. To facilitate modeling of this type, we develop a novel graph-based decoding framework that achieves state-of-the-art performance on the above datasets, and apply this method to two other benchmark SP tasks.

Modeling Semantic Plausibility by Injecting World Knowledge

Su Wang, Greg Durrett, and Katrin Erk

10:48–11:05

Distributional data tells us that a man can swallow candy, but not that a man can swallow a paintball, since this is never attested. However both are physically plausible events. This paper introduces the task of semantic plausibility: recognizing plausible but possibly novel events. We present a new crowdsourced dataset of semantic plausibility judgments of single events such as man swallow paintball. Simple models based on distributional representations perform poorly on this task, despite doing well on selection preference, but injecting manually elicited knowledge about entity properties provides a substantial performance boost. Our error analysis shows that our new dataset is a great testbed for semantic plausibility models: more sophisticated knowledge representation and propagation could address many of the remaining errors.

Neural Models of Factuality

Rachel Rudinger, Aaron Steven White, and Benjamin Van Durme

11:06–11:23

We present two neural models for event factuality prediction, which yield significant performance gains over previous models on three event factuality datasets: FactBank, UW, and MEANTIME. We also present a substantial expansion of the It Happened portion of the Universal Decompositional Semantics dataset, yielding the largest event factuality dataset to date. We report model results on this extended factuality dataset as well.

Industry: Machine Learning – Domain Adaptation

Empire D

Chair: Mohit Bansal

Selecting Machine-Translated Data for Quick Bootstrapping of a Natural Language Understanding System (INDUSTRY)

Judith Gaspers, Penny Karanasou, and Rajen Chatterjee

10:30–10:47

This paper investigates the use of Machine Translation (MT) to bootstrap a Natural Language Understanding (NLU) system for a new language for the use case of a large-scale voice-controlled device. The goal is to decrease the cost and time needed to get an annotated corpus for the new language, while still having a large enough coverage of user requests. Different methods of filtering MT data in order to keep utterances that improve NLU performance and language-specific post-processing methods are investigated. These methods are tested in a large-scale NLU task with translating around 10 millions training utterances from English to German. The results show a large improvement for using MT data over a grammar-based and over an in-house data collection baseline, while reducing the manual effort greatly. Both filtering and post-processing approaches improve results further.

Fast and Scalable Expansion of Natural Language Understanding Functionality for Intelligent Agents (INDUSTRY)

Anuj Kumar Goyal, Angeliki Metallinou, and Spyros Matsoukas

10:48–11:05

Fast expansion of natural language functionality of intelligent virtual agents is critical for achieving engaging and informative interactions. However, developing accurate models for new natural language domains is a time and data intensive process. We propose efficient deep neural network architectures that maximally reuse available resources through transfer learning. Our methods are applied for expanding the understanding capabilities of a popular commercial agent and are evaluated on hundreds of new domains, designed by internal or external developers. We demonstrate that our proposed methods significantly increase accuracy in low resource settings and enable rapid development of accurate models with less data.

Bag of Experts Architectures for Model Reuse in Conversational Language Understanding (INDUSTRY)

Rahul Jha, Alex Marin, Suvamsh Shivaprasad, and Imed Zitouni

11:06–11:23

Slot tagging, the task of detecting entities in input user utterances, is a key component of natural language understanding systems for personal digital assistants. Since each new domain requires a different set of slots, the annotation costs for labeling data for training slot tagging models increases rapidly as the number of domains grow. To tackle this, we describe Bag of Experts (BoE) architectures for model reuse for both LSTM and CRF based models. Extensive experimentation over a dataset of 10 domains drawn from data relevant to our commercial personal digital assistant shows that our BoE models outperform the baseline models with a statistically significant average margin of 5.06% in absolute F1-score when training with 2000 instances per domain, and achieve an even higher improvement of 12.16% when only 25% of the training data is used.

Session 6 Overview – Sunday, June 3, 2018

Track A	Track B	Track C	Track D	Track P
<i>Machine Learning 3</i> Empire A	<i>Social Media and Computational Social Science 2</i> Empire B	<i>Vision, Robotics and Other Grounding 2</i> Empire C	<i>Industry: Named Entity and Language Generation</i> Empire D	<i>Posters and Demos (10:30–12:00)</i> Elite Hall B
Deep Generative Model for Joint Alignment and Word Representation <i>M. Rios, W. Aziz, and K. Simaan</i>	Exploring the Role of Prior Beliefs for Argument Persuasion <i>E. Durmus and C. Cardie</i>	Defoiling Foiled Image Captions <i>P. S. Madhyastha, J. Wang, and L. Specia</i>	Multi-lingual Neural Title Generation for e-Commerce Browse Pages (INDUSTRY) <i>P. Mathur, N. Ueffing, and G. Leusch</i>	11:30
Evaluating the Stability of Embedding-based Word Similarities (TACL) <i>M. Antoniak and D. Mimno</i>	Inducing a Lexicon of Abusive Words – a Feature-Based Approach <i>M. Wiegand, J. Ruppenhofer, A. Schmidt, and C. Greenberg</i>	Pragmatically Informative Image Captioning with Character-Level Inference <i>R. Cohn-Gordon, N. Goodman, and C. Potts</i>	A Novel Approach to Part Name Discovery in Noisy Text (INDUSTRY) <i>N. B. Niraula, D. Whyatt, and A. Kao</i>	Information Extraction 3 / Tagging, Chunking, Syntax and Parsing 1 / Demos
Learning Word Embeddings for Low-resource Languages by PU Learning <i>C. Jiang, H.-F. Yu, C.-J. Hsieh, and K.-W. Chang</i>	Author Commitment and Social Power: Automatic Belief Tagging to Infer the Social Context of Interactions <i>V. Prabhakaran, P. Ganeshkumar, and O. Rambow</i>	Object Ordering with Bidirectional Matchings for Visual Reasoning <i>H. Tan and M. Bansal</i>	The Alexa Meaning Representation Language (INDUSTRY) <i>T. Kollar, D. Berry, L. Stuart, K. Owczarzak, T. Chung, L. Mathias, M. Kayser, B. Snow, and S. Matsoukas</i>	11:48 12:06

Parallel Session 6

Oral: Machine Learning 3

Empire A

Chair: Yulia Tsvetkov

Deep Generative Model for Joint Alignment and Word Representation

Miguel Rios, Wilker Aziz, and Khalil Simaan

11:30–11:47

This work exploits translation data as a source of semantically relevant learning signal for models of word representation. In particular, we exploit equivalence through translation as a form of distributional context and jointly learn how to embed and align with a deep generative model. Our EmbedAlign model embeds words in their complete observed context and learns by marginalisation of latent lexical alignments. Besides, it embeds words as posterior probability densities, rather than point estimates, which allows us to compare words in context using a measure of overlap between distributions (e.g. KL divergence). We investigate our model's performance on a range of lexical semantics tasks achieving competitive results on several standard benchmarks including natural language inference, paraphrasing, and text similarity.

Evaluating the Stability of Embedding-based Word Similarities (TACL)

Maria Antoniak and David Mimno

11:48–12:05

Word embeddings are increasingly being used as a tool to study word associations in specific corpora. However, it is unclear whether such embeddings reflect enduring properties of language or if they are sensitive to inconsequential variations in the source documents. We find that nearest-neighbor distances are highly sensitive to small changes in the training corpus for a variety of algorithms. For all methods, including specific documents in the training set can result in substantial variations. We show that these effects are more prominent for smaller training corpora. We recommend that users never rely on single embedding models for distance calculations, but rather average over multiple bootstrap samples, especially for small corpora.

Learning Word Embeddings for Low-resource Languages by PU Learning

Chao Jiang, Hsiang-Fu Yu, Cho-Jui Hsieh, and Kai-Wei Chang

12:06–12:23

Word embedding is a key component in many downstream applications in processing natural languages. Existing approaches often assume the existence of a large collection of text for learning effective word embedding. However, such a corpus may not be available for some low-resource languages. In this paper, we study how to effectively learn a word embedding model on a corpus with only a few million tokens. In such a situation, the co-occurrence matrix is sparse as the co-occurrences of many word pairs are unobserved. In contrast to existing approaches often only sample a few unobserved word pairs as negative samples, we argue that the zero entries in the co-occurrence matrix also provide valuable information. We then design a Positive-Unlabeled Learning (PU-Learning) approach to factorize the co-occurrence matrix and validate the proposed approaches in four different languages.

Oral: Social Media and Computational Social Science 2

Empire B

Chair: Nanyun Peng

Exploring the Role of Prior Beliefs for Argument Persuasion

Esin Durmus and Claire Cardie

11:30–11:47

Public debate forums provide a common platform for exchanging opinions on a topic of interest. While recent studies in natural language processing (NLP) have provided empirical evidence that the language of the debaters and their patterns of interaction play a key role in changing the mind of a reader, research in psychology has shown that prior beliefs can affect our interpretation of an argument and could therefore constitute a competing alternative explanation for resistance to changing one's stance. To study the actual effect of language use vs. prior beliefs on persuasion, we provide a new dataset and propose a controlled setting that takes into consideration two reader-level factors: political and religious ideology. We find that prior beliefs affected by these reader-level factors play a more important role than language use effects and argue that it is important to account for them in NLP studies of persuasion.

Inducing a Lexicon of Abusive Words – a Feature-Based Approach*Michael Wiegand, Josef Ruppenhofer, Anna Schmidt, and Clayton Greenberg*

11:48–12:05

We address the detection of abusive words. The task is to identify such words among a set of negative polar expressions. We propose novel features employing information from both corpora and lexical resources. These features are calibrated on a small manually annotated base lexicon which we use to produce a large lexicon. We show that the word-level information we learn cannot be equally derived from a large dataset of annotated microposts. We demonstrate the effectiveness of our (domain-independent) lexicon in the cross-domain detection of abusive microposts.

Author Commitment and Social Power: Automatic Belief Tagging to Infer the Social Context of Interactions*Vinodkumar Prabhakaran, Premkumar Ganeshkumar, and Owen Rambow*

12:06–12:23

Understanding how social power structures affect the way we interact with one another is of great interest to social scientists who want to answer fundamental questions about human behavior, as well as to computer scientists who want to build automatic methods to infer the social contexts of interactions. In this paper, we employ advancements in extra-propositional semantics extraction within NLP to study how author commitment reflects the social context of an interactions. Specifically, we investigate whether the level of commitment expressed by individuals in an organizational interaction reflects the hierarchical power structures they are part of. We find that subordinates use significantly more instances of non-commitment than superiors. More importantly, we also find that subordinates attribute propositions to other agents more often than superiors do — an aspect that has not been studied before. Finally, we show that enriching lexical features with commitment labels captures important distinctions in social meanings.

Oral: Vision, Robotics and Other Grounding 2

Empire C

Chair: Clare Voss

Defoiling Foiled Image Captions*Pranava Swaroop Madhyastha, Josiah Wang, and Lucia Specia*

11:30–11:47

We address the task of detecting foiled image captions, i.e. identifying whether a caption contains a word that has been deliberately replaced by a semantically similar word, thus rendering it inaccurate with respect to the image being described. Solving this problem should in principle require a fine-grained understanding of images to detect subtle perturbations in captions. In such contexts, encoding sufficiently descriptive image information becomes a key challenge. In this paper, we demonstrate that it is possible to solve this task using simple, interpretable yet powerful representations based on explicit object information over multilayer perceptron models. Our models achieve state-of-the-art performance on a recently published dataset, with scores exceeding those achieved by humans on the task. We also measure the upper-bound performance of our models using gold standard annotations. Our study and analysis reveals that the simpler model performs well even without image information, suggesting that the dataset contains strong linguistic bias.

Pragmatically Informative Image Captioning with Character-Level Inference*Reuben Cohn-Gordon, Noah Goodman, and Christopher Potts*

11:48–12:05

We combine a neural image captioner with a Rational Speech Acts (RSA) model to make a system that is pragmatically informative: its objective is to produce captions that are not merely true but also distinguish their inputs from similar images. Previous attempts to combine RSA with neural image captioning require an inference which normalizes over the entire set of possible utterances. This poses a serious problem of efficiency, previously solved by sampling a small subset of possible utterances. We instead solve this problem by implementing a version of RSA which operates at the level of characters ("a", "b", "c", ...) during the unrolling of the caption. We find that the utterance-level effect of referential captions can be obtained with only character-level decisions. Finally, we introduce an automatic method for testing the performance of pragmatic speaker models, and show that our model outperforms a non-pragmatic baseline as well as a word-level RSA captioner.

Object Ordering with Bidirectional Matchings for Visual Reasoning*Hao Tan and Mohit Bansal*

12:06–12:23

Visual reasoning with compositional natural language instructions, e.g., based on the newly-released Cornell Natural Language Visual Reasoning (NLVR) dataset, is a challenging task, where the model needs to have the ability to create an accurate mapping between the diverse phrases and the several objects placed in complex arrangements in the image. Further, this mapping needs to be processed to answer the question in the statement

given the ordering and relationship of the objects across three similar images. In this paper, we propose a novel end-to-end neural model for the NLVR task, where we first use joint bidirectional attention to build a two-way conditioning between the visual information and the language phrases. Next, we use an RL-based pointer network to sort and process the varying number of unordered objects (so as to match the order of the statement phrases) in each of the three images and then pool over the three decisions. Our model achieves strong improvements (of 4-6% absolute) over the state-of-the-art on both the structured representation and raw image versions of the dataset.

Industry: Named Entity and Language Generation

Empire D

Chair: Owen Rambow

Multi-lingual Neural Title Generation for e-Commerce Browse Pages (INDUSTRY)

Prashant Mathur, Nicola Ueffing, and Gregor Leusch

11:30–11:47

To provide better access of the inventory to buyers and better search engine optimization, e-Commerce websites are automatically generating millions of browse pages. A browse page consists of a set of slot name/value pairs within a given category, grouping multiple items which share some characteristics. These browse pages require a title describing the content of the page. Since the number of browse pages are huge, manual creation of these titles is infeasible. Previous statistical and neural approaches depend heavily on the availability of large amounts of data in a language. In this research, we apply sequence-to-sequence models to generate titles for high-resource as well as low-resource languages by leveraging transfer learning. We train these models on multi-lingual data, thereby creating one joint model which can generate titles in various different languages. Performance of the title generation system is evaluated on three different languages; English, German, and French, with a particular focus on low-resourced French language.

A Novel Approach to Part Name Discovery in Noisy Text (INDUSTRY)

Nobal Bikram Niraula, Daniel Whyatt, and Anne Kao

11:48–12:05

As a specialized example of information extraction, part name extraction is an area that presents unique challenges. Part names are typically multi-word terms longer than two words. There is little consistency in how terms are described in noisy free text, with variations spawned by typos, ad hoc abbreviations, acronyms, and incomplete names. This makes search and analyses of parts in these data extremely challenging. In this paper, we present our algorithm, PANDA (Pan Name Discovery Analytics), based on a unique method that exploits statistical, linguistic and machine learning techniques to discover part names in noisy text such as that in manufacturing quality documentation, supply chain management records, service communication logs, and maintenance reports. Experiments show that PANDA is scalable and outperforms existing techniques significantly.

The Alexa Meaning Representation Language (INDUSTRY)

Thomas Kollar, Danielle Berry, Lauren Stuart, Karolina Owczarzak, Tagyoung Chung, Lambert Mathias, Michael Kayser, Bradford Snow, and Spyros Matsoukas

12:06–12:23

This paper introduces a meaning representation for spoken language understanding. The Alexa meaning representation language (AMRL), unlike previous approaches, which factor spoken utterances into domains, provides a common representation for how people communicate in spoken language. AMRL is a rooted graph, links to a large-scale ontology, supports cross-domain queries, fine-grained types, complex utterances and composition. A spoken language dataset has been collected for Alexa, which contains 20k examples across eight domains. A version of this meaning representation was released to developers at a trade show in 2016.

Posters and Demos: Sessions 5/6

Posters: Information Extraction 3

Time: 10:30–12:00

Chair: Marine Carpuat

Location: Elite Hall B

A Bi-Model Based RNN Semantic Frame Parsing Model for Intent Detection and Slot Filling

Yu Wang, Yilin Shen, and Hongxia Jin

Intent detection and slot filling are two main tasks for building a spoken language understanding(SLU) system. Multiple deep learning based models have demonstrated good results on these tasks . The most effective algorithms are based on the structures of sequence to sequence models (or “encoder-decoder” models), and generate the intents and semantic tags either using separate models. Most of the previous studies, however, either treat the intent detection and slot filling as two separate parallel tasks, or use a sequence to sequence model to generate both semantic tags and intent. None of the approaches consider the cross-impact between the intent detection task and the slot filling task. In this paper, new Bi-model based RNN semantic frame parsing network structures are designed to perform the intent detection and slot filling tasks jointly, by considering their cross-impact to each other using two correlated bidirectional LSTMs (BLSTM). Our Bi-model structure with a decoder achieves state-of-art result on the benchmark ATIS data, with about 0.5% intent accuracy improvement and 0.9 % slot filling improvement.

A Comparison of Two Paraphrase Models for Taxonomy Augmentation

Vassilis Plachouras, Fabio Petroni, Timothy Nugent, and Jochen L. Leidner

Taxonomies are often used to look up the concepts they contain in text documents (for instance, to classify a document). The more comprehensive the taxonomy, the higher recall the application has that uses the taxonomy. In this paper, we explore automatic taxonomy augmentation with paraphrases. We compare two state-of-the-art paraphrase models based on Moses, a statistical Machine Translation system, and a sequence-to-sequence neural network, trained on a paraphrase datasets with respect to their abilities to add novel nodes to an existing taxonomy from the risk domain. We conduct component-based and task-based evaluations. Our results show that paraphrasing is a viable method to enrich a taxonomy with more terms, and that Moses consistently outperforms the sequence-to-sequence neural model. To the best of our knowledge, this is the first approach to augment taxonomies with paraphrases.

A Laypeople Study on Terminology Identification across Domains and Task Definitions

Anna Häfty and Sabine Schulte im Walde

This paper introduces a new dataset of term annotation. Given that even experts vary significantly in their understanding of termhood, and that term identification is mostly performed as a binary task, we offer a novel perspective to explore the common, natural understanding of what constitutes a term: Laypeople annotate single-word and multi-word terms, across four domains and across four task definitions. Analyses based on inter-annotator agreement offer insights into differences in term specificity, term granularity and subtermhood.

A Novel Embedding Model for Knowledge Base Completion Based on Convolutional Neural Network

Dai Quoc Nguyen, Tu Dinh Nguyen, Dat Quoc Nguyen, and Dinh Phung

In this paper, we propose a novel embedding model, named ConvKB, for knowledge base completion. Our model ConvKB advances state-of-the-art models by employing a convolutional neural network, so that it can capture global relationships and transitional characteristics between entities and relations in knowledge bases. In ConvKB, each triple (head entity, relation, tail entity) is represented as a 3-column matrix where each column vector represents a triple element. This 3-column matrix is then fed to a convolution layer where multiple filters are operated on the matrix to generate different feature maps. These feature maps are then concatenated into a single feature vector representing the input triple. The feature vector is multiplied with a weight vector via a dot product to return a score. This score is then used to predict whether the triple is valid or not. Experiments show that ConvKB achieves better link prediction performance than previous state-of-the-art embedding models on two benchmark datasets WN18RR and FB15k-237.

Accurate Text-Enhanced Knowledge Graph Representation Learning*Bo An, Bo Chen, Xianpei Han, and Le Sun*

Previous representation learning techniques for knowledge graph representation usually represent the same entity or relation in different triples with the same representation, without considering the ambiguity of relations and entities. To appropriately handle the semantic variety of entities/relations in distinct triples, we propose an accurate text-enhanced knowledge graph representation learning method, which can represent a relation/entity with different representations in different triples by exploiting additional textual information. Specifically, our method enhances representations by exploiting the entity descriptions and triple-specific relation mention. And a mutual attention mechanism between relation mention and entity description is proposed to learn more accurate textual representations for further improving knowledge graph representation. Experimental results show that our method achieves the state-of-the-art performance on both link prediction and triple classification tasks, and significantly outperforms previous text-enhanced knowledge representation models.

Acquisition of Phrase Correspondences Using Natural Deduction Proofs*Hitomi Yanaka, Koji Mineshima, Pascual Martínez-Gómez, and Daisuke Bekki*

How to identify, extract, and use phrasal knowledge is a crucial problem for the task of Recognizing Textual Entailment (RTE). To solve this problem, we propose a method for detecting paraphrases via natural deduction proofs of semantic relations between sentence pairs. Our solution relies on a graph reformulation of partial variable unifications and an algorithm that induces subgraph alignments between meaning representations. Experiments show that our method can automatically detect various paraphrases that are absent from existing paraphrase databases. In addition, the detection of paraphrases using proof information improves the accuracy of RTE tasks.

Automatic Stance Detection Using End-to-End Memory Networks*Mitra Mohtarami, Ramy Baly, James Glass, Preslav Nakov, Lluís Márquez, and Alessandro Moschitti*

We present an effective end-to-end memory network model that jointly (i) predicts whether a given document can be considered as relevant evidence for a given claim, and (ii) extracts snippets of evidence that can be used to reason about the factuality of the target claim. Our model combines the advantages of convolutional and recurrent neural networks as part of a memory network. We further introduce a similarity matrix at the inference level of the memory network in order to extract snippets of evidence for input claims more accurately. Our experiments on a public benchmark dataset, FakeNewsChallenge, demonstrate the effectiveness of our approach.

Collective Entity Disambiguation with Structured Gradient Tree Boosting*Yi Yang, Ozan Irsoy, and Kazi Shefaet Rahman*

We present a gradient-tree-boosting-based structured learning model for jointly disambiguating named entities in a document. Gradient tree boosting is a widely used machine learning algorithm that underlies many top-performing natural language processing systems. Surprisingly, most works limit the use of gradient tree boosting as a tool for regular classification or regression problems, despite the structured nature of language. To the best of our knowledge, our work is the first one that employs the structured gradient tree boosting (SGTB) algorithm for collective entity disambiguation. By defining global features over previous disambiguation decisions and jointly modeling them with local features, our system is able to produce globally optimized entity assignments for mentions in a document. Exact inference is prohibitively expensive for our globally normalized model. To solve this problem, we propose Bidirectional Beam Search with Gold path (BiBSG), an approximate inference algorithm that is a variant of the standard beam search algorithm. BiBSG makes use of global information from both past and future to perform better local search. Experiments on standard benchmark datasets show that SGTB significantly improves upon published results. Specifically, SGTB outperforms the previous state-of-the-art neural system by near 1% absolute accuracy on the popular AIDA-CoNLL dataset.

Cross-language Article Linking Using Cross-Encyclopedia Entity Embedding*Chun-Kai Wu and Richard Tzong-Han Tsai*

Cross-language article linking (CLAL) is the task of finding corresponding article pairs of different languages across encyclopedias. This task is a difficult disambiguation problem in which one article must be selected among several candidate articles with similar titles and contents. Existing works focus on engineering text-based or link-based features for this task, which is a time-consuming job, and some of these features are only applicable within the same encyclopedia. In this paper, we address these problems by proposing cross-encyclopedia entity embedding. Unlike other works, our proposed method does not rely on known cross-language pairs. We apply our method to CLAL between English Wikipedia and Chinese Baidu Baike. Our

features improve performance relative to the baseline by 29.62%. Tested 30 times, our system achieved an average improvement of 2.76% over the current best system (26.86% over baseline), a statistically significant result.

DeepAlignment: Unsupervised Ontology Matching with Refined Word Vectors

Prodromos Kolylakis, Alexandros Kalousis, and Dimitris Kiritidis

Ontologies compartmentalize types and relations in a target domain and provide the semantic backbone needed for a plethora of practical applications. Very often different ontologies are developed independently for the same domain. Such “parallel” ontologies raise the need for a process that will establish alignments between their entities in order to unify and extend the existing knowledge. In this work, we present a novel entity alignment method which we dub DeepAlignment. DeepAlignment refines pre-trained word vectors aiming at deriving ontological entity descriptions which are tailored to the ontology matching task. The absence of explicit information relevant to the ontology matching task during the refinement process makes DeepAlignment completely unsupervised. We empirically evaluate our method using standard ontology matching benchmarks. We present significant performance improvements over the current state-of-the-art, demonstrating the advantages that representation learning techniques bring to ontology matching.

Efficient Sequence Learning with Group Recurrent Networks

Fei Gao, Lijun Wu, Li Zhao, Tao Qin, Xueqi Cheng, and Tie-Yan Liu

Recurrent neural networks have achieved state-of-the-art results in many artificial intelligence tasks, such as language modeling, neural machine translation, speech recognition and so on. One of the key factors to these successes is big models. However, training such big models usually takes days or even weeks of time even if using tens of GPU cards. In this paper, we propose an efficient architecture to improve the efficiency of such RNN model training, which adopts the group strategy for recurrent layers, while exploiting the representation rearrangement strategy between layers as well as time steps. To demonstrate the advantages of our models, we conduct experiments on several datasets and tasks. The results show that our architecture achieves comparable or better accuracy comparing with baselines, with a much smaller number of parameters and at a much lower computational cost.

FEVER: a Large-scale Dataset for Fact Extraction and VERification

James Thorne, Andreas Vlachos, Christos Christodoulopoulos, and Arpit Mittal

In this paper we introduce a new publicly available dataset for verification against textual sources, FEVER: Fact Extraction and VERification. It consists of 185,445 claims generated by altering sentences extracted from Wikipedia and subsequently verified without knowledge of the sentence they were derived from. The claims are classified as Supported, Refuted or NotEnoughInfo by annotators achieving 0.6841 in Fleiss kappa. For the first two classes, the annotators also recorded the sentence(s) forming the necessary evidence for their judgment. To characterize the challenge of the dataset presented, we develop a pipeline approach and compare it to suitably designed oracles. The best accuracy we achieve on labeling a claim accompanied by the correct evidence is 31.87%, while if we ignore the evidence we achieve 50.91%. Thus we believe that FEVER is a challenging testbed that will help stimulate progress on claim verification against textual sources.

Global Relation Embedding for Relation Extraction

Yu Su, Honglei Liu, Semih Yavuz, Izzeddin Gur, Huan Sun, and Xifeng Yan

We study the problem of textual relation embedding with distant supervision. To combat the wrong labeling problem of distant supervision, we propose to embed textual relations with global statistics of relations, i.e., the co-occurrence statistics of textual and knowledge base relations collected from the entire corpus. This approach turns out to be more robust to the training noise introduced by distant supervision. On a popular relation extraction dataset, we show that the learned textual relation embedding can be used to augment existing relation extraction models and significantly improve their performance. Most remarkably, for the top 1,000 relational facts discovered by the best existing model, the precision can be improved from 83.9% to 89.3%.

Identifying the Most Dominant Event in a News Article by Mining Event Coreference Relations

Prafulla Kumar Choubey, Kaushik Raju, and Ruihong Huang

Identifying the most dominant and central event of a document, which governs and connects other foreground and background events in the document, is useful for many applications, such as text summarization, storyline generation and text segmentation. We observed that the central event of a document usually has many coreferential event mentions that are scattered throughout the document for enabling a smooth transition of subtopics. Our empirical experiments, using gold event coreference relations, have shown that the central event of a doc-

ument can be well identified by mining properties of event coreference chains. But the performance drops when switching to system predicted event coreference relations. In addition, we found that the central event can be more accurately identified by further considering the number of sub-events as well as the realis status of an event.

Implicit Argument Prediction with Event Knowledge

Pengxiang Cheng and Katrin Erk

Implicit arguments are not syntactically connected to their predicates, and are therefore hard to extract. Previous work has used models with large numbers of features, evaluated on very small datasets. We propose to train models for implicit argument prediction on a simple cloze task, for which data can be generated automatically at scale. This allows us to use a neural model, which draws on narrative coherence and entity salience for predictions. We show that our model has superior performance on both synthetic and natural data.

Improve Neural Entity Recognition via Multi-Task Data Selection and Constrained Decoding

Huasha Zhao, Yi Yang, Qiong Zhang, and Luo Si

Entity recognition is a widely benchmarked task in natural language processing due to its massive applications. The state-of-the-art solution applies a neural architecture named BiLSTM-CRF to model the language sequences. In this paper, we propose an entity recognition system that improves this neural architecture with two novel techniques. The first technique is Multi-Task Data Selection, which ensures the consistency of data distribution and labeling guidelines between source and target datasets. The other one is constrained decoding using knowledge base. The decoder of the model operates at the document level, and leverages global and external information sources to further improve performance. Extensive experiments have been conducted to show the advantages of each technique. Our system achieves state-of-the-art results on the English entity recognition task in KBP 2017 official evaluation, and it also yields very strong results in other languages.

Improving Temporal Relation Extraction with a Globally Acquired Statistical Resource

Qiang Ning, Hao Wu, Haoruo Peng, and Dan Roth

Extracting temporal relations (before, after, overlapping, etc.) is a key aspect of understanding events described in natural language. We argue that this task would gain from the availability of a resource that provides prior knowledge in the form of the temporal order that events usually follow. This paper develops such a resource – a probabilistic knowledge base acquired in the news domain – by extracting temporal relations between events from the New York Times (NYT) articles over a 20-year span (1987–2007). We show that existing temporal extraction systems can be improved via this resource. As a byproduct, we also show that interesting statistics can be retrieved from this resource, which can potentially benefit other time-aware tasks. The proposed system and resource are both publicly available.

Keep Your Bearings: Lightly-Supervised Information Extraction with Ladder Networks That Avoids Semantic Drift

Ajay Nagesh and Mihai Surdeanu

We propose a novel approach to semi-supervised learning for information extraction that uses ladder networks (Rasmus et al., 2015). In particular, we focus on the task of named entity classification, defined as identifying the correct label (e.g., person or organization name) of an entity mention in a given context. Our approach is simple, efficient and has the benefit of being robust to semantic drift, a dominant problem in most semi-supervised learning systems. We empirically demonstrate the superior performance of our system compared to the state-of-the-art on two standard datasets for named entity classification. We obtain between 62% and 200% improvement over the state-of-art baseline on these two datasets.

Multimodal Named Entity Recognition for Short Social Media Posts

Seungwhan Moon, Leonardo Neves, and Vitor Carvalho

We introduce a new task called Multimodal Named Entity Recognition (MNER) for noisy user-generated data such as tweets or Snapchat captions, which comprise short text with accompanying images. These social media posts often come in inconsistent or incomplete syntax and lexical notations with very limited surrounding textual contexts, bringing significant challenges for NER. To this end, we create a new dataset for MNER called SnapCaptions (Snapchat image-caption pairs submitted to public and crowd-sourced stories with fully annotated named entities). We then build upon the state-of-the-art Bi-LSTM word/character based NER models with 1) a deep image network which incorporates relevant visual context to augment textual information, and 2) a generic *modality-attention* module which learns to attenuate irrelevant modalities while amplifying the most informative ones to extract contexts from, adaptive to each sample and token. The proposed MNER model with modality attention significantly outperforms the state-of-the-art text-only NER models by success-

fully leveraging provided visual contexts, opening up potential applications of MNER on myriads of social media platforms.

Nested Named Entity Recognition Revisited

Arzoo Katiyar and Claire Cardie

We propose a novel recurrent neural network-based approach to simultaneously handle nested named entity recognition and nested entity mention detection. The model learns a hypergraph representation for nested entities using features extracted from a recurrent neural network. In evaluations on three standard data sets, we show that our approach significantly outperforms existing state-of-the-art methods, which are feature-based. The approach is also efficient: it operates linearly in the number of tokens and the number of possible output labels at any token. Finally, we present an extension of our model that jointly learns the head of each entity mention.

Semi-Supervised Event Extraction with Paraphrase Clusters

James Ferguson, Colin Lockard, Daniel Weld, and Hannaneh Hajishirzi

Supervised event extraction systems are limited in their accuracy due to the lack of available training data. We present a method for self-training event extraction systems by bootstrapping additional training data. This is done by taking advantage of the occurrence of multiple mentions of the same event instances across newswire articles from multiple sources. If our system can make a high-confidence extraction of some mentions in such a cluster, it can then acquire diverse training examples by adding the other mentions as well. Our experiments show significant performance improvements on multiple event extractors over ACE 2005 and TAC-KBP 2015 datasets.

Simultaneously Self-Attending to All Mentions for Full-Abstract Biological Relation Extraction

Patrick Verga, Emma Strubell, and Andrew McCallum

Most work in relation extraction forms a prediction by looking at a short span of text within a single sentence containing a single entity pair mention. This approach often does not consider interactions across mentions, requires redundant computation for each mention pair, and ignores relationships expressed across sentence boundaries. These problems are exacerbated by the document- (rather than sentence-) level annotation common in biological text. In response, we propose a model which simultaneously predicts relationships between all mention pairs in a document. We form pairwise predictions over entire paper abstracts using an efficient self-attention encoder. All-pairs mention scores allow us to perform multi-instance learning by aggregating over mentions to form entity pair representations. We further adapt to settings without mention-level annotation by jointly training to predict named entities and adding a corpus of weakly labeled data. In experiments on two Biocreative benchmark datasets, we achieve state of the art performance on the Biocreative V Chemical Disease Relation dataset for models without external KB resources. We also introduce a new dataset an order of magnitude larger than existing human-annotated biological information extraction datasets and more accurate than distantly supervised alternatives.

Structure Regularized Neural Network for Entity Relation Classification for Chinese Literature Text

Ji Wen, Xu Sun, Xuancheng Ren, and Qi Su

Relation classification is an important semantic processing task in the field of natural language processing. In this paper, we propose the task of relation classification for Chinese literature text. A new dataset of Chinese literature text is constructed to facilitate the study in this task. We present a novel model, named Structure Regularized Bidirectional Recurrent Convolutional Neural Network (SR-BRCNN), to identify the relation between entities. The proposed model learns relation representations along the shortest dependency path (SDP) extracted from the structure regularized dependency tree, which has the benefits of reducing the complexity of the whole model. Experimental results show that the proposed method significantly improves the F1 score by 10.3, and outperforms the state-of-the-art approaches on Chinese literature text.

Supervised Open Information Extraction

Gabriel Stanovsky, Julian Michael, Luke Zettlemoyer, and Ido Dagan

We present data and methods that enable a supervised learning approach to Open Information Extraction (Open IE). Central to the approach is a novel formulation of Open IE as a sequence tagging problem, addressing challenges such as encoding multiple extractions for a predicate. We also develop a bi-LSTM transducer, extending recent deep Semantic Role Labeling models to extract Open IE tuples and provide confidence scores for tuning their precision-recall tradeoff. Furthermore, we show that the recently released Question-Answer Meaning Representation dataset can be automatically converted into an Open IE corpus which significantly

increases the amount of available training data. Our supervised model outperforms the existing state-of-the-art Open IE systems on benchmark datasets.

Syntactic Patterns Improve Information Extraction for Medical Search

Roma Patel, Yinfei Yang, Iain Marshall, Ani Nenkova, and Byron Wallace

Medical professionals search the published literature by specifying the type of patients, the medical intervention(s) and the outcome measure(s) of interest. In this paper we demonstrate how features encoding syntactic patterns improve the performance of state-of-the-art sequence tagging models (both neural and linear) for information extraction of these medically relevant categories. We present an analysis of the type of patterns exploited and of the semantic space induced for these, i.e., the distributed representations learned for identified multi-token patterns. We show that these learned representations differ substantially from those of the constituent unigrams, suggesting that the patterns capture contextual information that is otherwise lost.

Syntactically Aware Neural Architectures for Definition Extraction

Luis Espinosa Anke and Steven Schockaert

Automatically identifying definitional knowledge in text corpora (Definition Extraction or DE) is an important task with direct applications in, among others, Automatic Glossary Generation, Taxonomy Learning, Question Answering and Semantic Search. It is generally cast as a binary classification problem between definitional and non-definitional sentences. In this paper we present a set of neural architectures combining Convolutional and Recurrent Neural Networks, which are further enriched by incorporating linguistic information via syntactic dependencies. Our experimental results in the task of sentence classification, on two benchmarking DE datasets (one generic, one domain-specific), show that these models obtain consistent state of the art results. Furthermore, we demonstrate that models trained on clean Wikipedia-like definitions can successfully be applied to more noisy domain-specific corpora.

Read and Comprehend by Gated-Attention Reader with More Belief (SRW)

Haohui Deng and Yik-Cheung Tam

Gated-Attention (GA) Reader has been effective for reading comprehension. GA Reader makes two assumptions: (1) a uni-directional attention that uses an input query to gate token encodings of a document; (2) encoding at the cloze position of an input query is considered for answer prediction. In this paper, we propose Collaborative Gating (CG) and Self-Belief Aggregation (SBA) to address the above assumptions respectively. In CG, we first use an input document to gate token encodings of an input query so that the influence of irrelevant query tokens may be reduced. Then the filtered query is used to gate token encodings of an document in a collaborative fashion. In SBA, we conjecture that query tokens other than the cloze token may be informative for answer prediction. We apply self-attention to link the cloze token with other tokens in a query so that the importance of query tokens with respect to the cloze position are weighted. Then their evidences are weighted, propagated and aggregated for better reading comprehension. Experiments show that our approaches advance the state-of-the-art results in CNN, Daily Mail, and Who Did What public test sets.

Posters: Tagging, Chunking, Syntax and Parsing 1

Time: 10:30–12:00

Location: Elite Hall B

Chair: Marine Carpuat

A Dynamic Oracle for Linear-Time 2-Planar Dependency Parsing

Daniel Fernández-González and Carlos Gómez-Rodríguez

We propose an efficient dynamic oracle for training the 2-Planar transition-based parser, a linear-time parser with over 99% coverage on non-projective syntactic corpora. This novel approach outperforms the static training strategy in the vast majority of languages tested and scored better on most datasets than the arc-hybrid parser enhanced with the Swap transition, which can handle unrestricted non-projectivity.

Are Automatic Methods for Cognate Detection Good Enough for Phylogenetic Reconstruction in Historical Linguistics?

Taraka Rama, Johann-Mattis List, Johannes Wahle, and Gerhard Jäger

We evaluate the performance of state-of-the-art algorithms for automatic cognate detection by comparing how useful automatically inferred cognates are for the task of phylogenetic inference compared to classical manually annotated cognate sets. Our findings suggest that phylogenies inferred from automated cognate sets come close to phylogenies inferred from expert-annotated ones, although on average, the latter are still superior. We conclude that future work on phylogenetic reconstruction can profit much from automatic cognate detection.

Especially where scholars are merely interested in exploring the bigger picture of a language family's phylogeny, algorithms for automatic cognate detection are a useful complement for current research on language phylogenies.

Automatically Selecting the Best Dependency Annotation Design with Dynamic Oracles
Guillaume Wisniewski, Ophélie Lacroix, and François Yvon

This work introduces a new strategy to compare the numerous conventions that have been proposed over the years for expressing dependency structures and discover the one for which a parser will achieve the highest parsing performance. Instead of associating each sentence in the training set with a single gold reference we propose to consider a set of references encoding alternative syntactic representations. Training a parser with a dynamic oracle will then automatically select among all alternatives the reference that will be predicted with the highest accuracy. Experiments on the UD corpora show the validity of this approach.

Consistent CCG Parsing over Multiple Sentences for Improved Logical Reasoning
Masashi Yoshikawa, Koji Mineshima, Hiroshi Noji, and Daisuke Bekki

In formal logic-based approaches to Recognizing Textual Entailment (RTE), a Combinatory Categorial Grammar (CCG) parser is used to parse input premises and hypotheses to obtain their logical formulas. Here, it is important that the parser processes the sentences consistently; failing to recognize the similar syntactic structure results in inconsistent predicate argument structures among them, in which case the succeeding theorem proving is doomed to failure. In this work, we present a simple method to extend an existing CCG parser to parse a set of sentences consistently, which is achieved with an inter-sentence modeling with Markov Random Fields (MRF). When combined with existing logic-based systems, our method always shows improvement in the RTE experiments on English and Japanese languages.

Embedding Syntax and Semantics of Prepositions via Tensor Decomposition
Hongyu Gong, Suma Bhat, and Pramod Viswanath

Prepositions are among the most frequent words in English and play complex roles in the syntax and semantics of sentences. Not surprisingly, they pose well-known difficulties in automatic processing of sentences (prepositional attachment ambiguities and idiosyncratic uses in phrases). Existing methods on preposition representation treat prepositions no different from content words (e.g., word2vec and GloVe). In addition, recent studies aiming at solving prepositional attachment and preposition selection problems depend heavily on external linguistic resources and use dataset-specific word representations. In this paper we use *word-triple* counts (one of the triples being a preposition) to capture a preposition's interaction with its attachment and complement. We then derive preposition embeddings via tensor decomposition on a large unlabeled corpus. We reveal a new geometry involving Hadamard products and empirically demonstrate its utility in paraphrasing phrasal verbs. Furthermore, our preposition embeddings are used as simple features in two challenging downstream tasks: preposition selection and prepositional attachment disambiguation. We achieve results comparable to or better than the state-of-the-art on multiple standardized datasets.

Exploiting Dynamic Oracles to Train Projective Dependency Parsers on Non-Projective Trees

Lauriane Aufrant, Guillaume Wisniewski, and François Yvon

Because the most common transition systems are projective, training a transition-based dependency parser often implies to either ignore or rewrite the non-projective training examples, which has an adverse impact on accuracy. In this work, we propose a simple modification of dynamic oracles, which enables the use of non-projective data when training projective parsers. Evaluation on 73 treebanks shows that our method achieves significant gains (+2 to +7 UAS for the most non-projective languages) and consistently outperforms traditional projectivization and pseudo-projectivization approaches.

From Phonology to Syntax: Unsupervised Linguistic Typology at Different Levels with Language Embeddings

Johannes Bjerva and Isabelle Augenstein

A core part of linguistic typology is the classification of languages according to linguistic properties, such as those detailed in the World Atlas of Language Structure (WALS). Doing this manually is prohibitively time-consuming, which is in part evidenced by the fact that only 100 out of over 7,000 languages spoken in the world are fully covered in WALS. We learn distributed language representations, which can be used to predict typological properties on a massively multilingual scale. Additionally, quantitative and qualitative analyses of these language embeddings can tell us how language similarities are encoded in NLP models for tasks at different typological levels. The representations are learned in an unsupervised manner alongside tasks at three

typological levels: phonology (grapheme-to-phoneme prediction, and phoneme reconstruction), morphology (morphological inflection), and syntax (part-of-speech tagging). We consider more than 800 languages and find significant differences in the language representations encoded, depending on the target task. For instance, although Norwegian Bokmål and Danish are typologically close to one another, they are phonologically distant, which is reflected in their language embeddings growing relatively distant in a phonological task. We are also able to predict typological features in WALS with high accuracies, even for unseen language families.

Improving Coverage and Runtime Complexity for Exact Inference in Non-Projective Transition-Based Dependency Parsers

Tianze Shi, Carlos Gómez-Rodríguez, and Lillian Lee

We generalize Cohen, Gómez-Rodríguez, and Satta’s (2011) parser to a family of non-projective transition-based dependency parsers allowing polynomial-time exact inference. This includes novel parsers with better coverage than Cohen et al. (2011), and even a variant that reduces time complexity to $O(n^6)$, improving over the known bounds in exact inference for non-projective transition-based parsing. We hope that this piece of theoretical work inspires design of novel transition systems with better coverage and better run-time guarantees.

Monte Carlo Syntax Marginals for Exploring and Using Dependency Parses

Katherine Keith, Su Lin Blodgett, and Brendan O’Connor

Dependency parsing research, which has made significant gains in recent years, typically focuses on improving the accuracy of single-tree predictions. However, ambiguity is inherent to natural language syntax, and communicating such ambiguity is important for error analysis and better-informed downstream applications. In this work, we propose a transition sampling algorithm to sample from the full joint distribution of parse trees defined by a transition-based parsing model, and demonstrate the use of the samples in probabilistic dependency analysis. First, we define the new task of dependency path prediction, inferring syntactic substructures over part of a sentence, and provide the first analysis of performance on this task. Second, we demonstrate the usefulness of our Monte Carlo syntax marginal method for parser error analysis and calibration. Finally, we use this method to propagate parse uncertainty to two downstream information extraction applications: identifying persons killed by police and semantic role assignment.

Neural Particle Smoothing for Sampling from Conditional Sequence Models

Chu-Cheng Lin and Jason Eisner

We introduce neural particle smoothing, a sequential Monte Carlo method for sampling annotations of an input string from a given probability model. In contrast to conventional particle filtering algorithms, we train a proposal distribution that looks ahead to the end of the input string by means of a right-to-left LSTM. We demonstrate that this innovation can improve the quality of the sample. To motivate our formal choices, we explain how neural transduction models and our sampler can be viewed as low-dimensional but nonlinear approximations to working with HMMs over very large state spaces.

Neural Syntactic Generative Models with Exact Marginalization

Jan Buys and Phil Blunsom

We present neural syntactic generative models with exact marginalization that support both dependency parsing and language modeling. Exact marginalization is made tractable through dynamic programming over shift-reduce parsing and minimal RNN-based feature sets. Our algorithms complement previous approaches by supporting batched training and enabling online computation of next word probabilities. For supervised dependency parsing, our model achieves a state-of-the-art result among generative approaches. We also report empirical results on unsupervised syntactic models and their role in language modeling. We find that our model formulation of latent dependencies with exact marginalization do not lead to better intrinsic language modeling performance than vanilla RNNs, and that parsing accuracy is not correlated with language modeling perplexity in stack-based models.

Noise-Robust Morphological Disambiguation for Dialectal Arabic

Nasser Zalmout, Alexander Erdmann, and Nizar Habash

User-generated text tends to be noisy with many lexical and orthographic inconsistencies, making natural language processing (NLP) tasks more challenging. The challenging nature of noisy text processing is exacerbated for dialectal content, where in addition to spelling and lexical differences, dialectal text is characterized with morpho-syntactic and phonetic variations. These issues increase sparsity in NLP models and reduce accuracy. We present a neural morphological tagging and disambiguation model for Egyptian Arabic, with various extensions to handle noisy and inconsistent content. Our models achieve about 5% relative error reduction

(1.1% absolute improvement) for full morphological analysis, and around 22% relative error reduction (1.8% absolute improvement) for part-of-speech tagging, over a state-of-the-art baseline.

Parsing Tweets into Universal Dependencies

Yijia Liu, Yi Zhu, Wanxiang Che, Bing Qin, Nathan Schneider, and Noah A. Smith

We study the problem of analyzing tweets with universal dependencies (UD). We extend the UD guidelines to cover special constructions in tweets that affect tokenization, part-of-speech tagging, and labeled dependencies. Using the extended guidelines, we create a new tweet treebank for English (Tweebank v2) that is four times larger than the (unlabeled) Tweebank v1 introduced by Kong et al. (2014). We characterize the disagreements between our annotators and show that it is challenging to deliver consistent annotation due to ambiguity in understanding and explaining tweets. Nonetheless, using the new treebank, we build a pipeline system to parse raw tweets into UD. To overcome the annotation noise without sacrificing computational efficiency, we propose a new method to distill an ensemble of 20 transition-based parsers into a single one. Our parser achieves an improvement of 2.2 in LAS over the un-ensembled baseline and outperforms parsers that are state-of-the-art on other treebanks in both accuracy and speed.

Robust Multilingual Part-of-Speech Tagging via Adversarial Training

Michihiro Yasunaga, Jungsik Kasai, and Dragomir Radev

Adversarial training (AT) is a powerful regularization method for neural networks, aiming to achieve robustness to input perturbations. Yet, the specific effects of the robustness obtained from AT are still unclear in the context of natural language processing. In this paper, we propose and analyze a neural POS tagging model that exploits AT. In our experiments on the Penn Treebank WSJ corpus and the Universal Dependencies (UD) dataset (27 languages), we find that AT not only improves the overall tagging accuracy, but also 1) prevents over-fitting well in low resource languages and 2) boosts tagging accuracy for rare / unseen words. We also demonstrate that 3) the improved tagging performance by AT contributes to the downstream task of dependency parsing, and that 4) AT helps the model to learn cleaner word representations. 5) The proposed AT model is generally effective in different sequence labeling tasks. These positive results motivate further use of AT for natural language tasks.

Towards a Variability Measure for Multiword Expressions

Caroline Pasquer, Agata Savary, Jean-Yves Antoine, and Carlos Ramisch

One of the most outstanding properties of multiword expressions (MWEs), especially verbal ones (VMWEs), important both in theoretical models and applications, is their idiosyncratic variability. Some MWEs are always continuous, while some others admit certain types of insertions. Components of some MWEs are rarely or never modified, while some others admit either specific or unrestricted modification. This unpredictable variability profile of MWEs hinders modeling and processing them as “words-with-spaces” on the one hand, and as regular syntactic structures on the other hand. Since variability of MWEs is a matter of scale rather than a binary property, we propose a 2-dimensional language-independent measure of variability dedicated to verbal MWEs based on syntactic and discontinuity-related clues. We assess its relevance with respect to a linguistic benchmark and its utility for the tasks of VMWE classification and variant identification on a French corpus.

Universal Dependency Parsing for Hindi-English Code-Switching

Irshad Bhat, Riyaz A. Bhat, Manish Shrivastava, and Dipti Sharma

Code-switching is a phenomenon of mixing grammatical structures of two or more languages under varied social constraints. The code-switching data differ so radically from the benchmark corpora used in NLP community that the application of standard technologies to these data degrades their performance sharply. Unlike standard corpora, these data often need to go through additional processes such as language identification, normalization and/or back-transliteration for their efficient processing. In this paper, we investigate these indispensable processes and other problems associated with syntactic parsing of code-switching data and propose methods to mitigate their effects. In particular, we study dependency parsing of code-switching data of Hindi and English multilingual speakers from Twitter. We present a treebank of Hindi-English code-switching tweets under Universal Dependencies scheme and propose a neural stacking model for parsing that efficiently leverages the part-of-speech tag and syntactic tree annotations in the code-switching treebank and the pre-existing Hindi and English treebanks. We also present normalization and back-transliteration models with a decoding process tailored for code-switching data. Results show that our neural stacking parser is 1.5% LAS points better than the augmented parsing model and 3.8% LAS points better than the one which uses first-best normalization and/or back-transliteration.

What's Going On in Neural Constituency Parsers? An Analysis

David Gaddy, Mitchell Stern, and Dan Klein

A number of differences have emerged between modern and classic approaches to constituency parsing in recent years, with structural components like grammars and feature-rich lexicons becoming less central while recurrent neural network representations rise in popularity. The goal of this work is to analyze the extent to which information provided directly by the model structure in classical systems is still being captured by neural methods. To this end, we propose a high-performance neural model (92.08 F1 on PTB) that is representative of recent work and perform a series of investigative experiments. We find that our model implicitly learns to encode much of the same information that was explicitly provided by grammars and lexicons in the past, indicating that this scaffolding can largely be subsumed by powerful general-purpose neural machinery.

Demos

Time: 10:30–12:00

Location: Elite Hall B

ELISA-EDL: A Cross-lingual Entity Extraction, Linking and Localization System

Boliang Zhang, Ying Lin, Xiaoman Pan, Di Lu, Jonathan May, Kevin Knight, and Heng Ji

We demonstrate ELISA-EDL, a state-of-the-art re-trainable system to extract entity mentions from low-resource languages, link them to external English knowledge bases, and visualize locations related to disaster topics on a world heatmap. We make all of our data sets, resources and system training and testing APIs publicly available for research purpose.

Entity Resolution and Location Disambiguation in the Ancient Hindu Temples Domain using Web Data

Ayush Maheshwari, Vishwajeet Kumar, Ganesh Ramakrishnan, and J. Saketha Nath

We present a system for resolving entities and disambiguating locations based on publicly available web data in the domain of ancient Hindu Temples. Scarce, unstructured information poses a challenge to Entity Resolution(ER) and snippet ranking. Additionally, because the same set of entities may be associated with multiple locations, Location Disambiguation(LD) is a problem. The mentions and descriptions of temples 1 exist in the order of hundreds of thousands, with such data generated by various users in various forms such as text (Wikipedia pages), videos (YouTube videos), blogs, etc. We demonstrate an integrated approach using a combination of grammar rules for parsing and unsupervised (clustering) algorithms to resolve entity and locations with high confidence. A demo of our system is accessible at tinyurl.com/templedemos 2 . Our system is open source and available on GitHub 3 .

Madly Ambiguous: A Game for Learning about Structural Ambiguity and Why It's Hard for Computers

Ajda Gokcen, Ethan Hill, and Michael White

Madly Ambiguous is an open source, online game aimed at teaching audiences of all ages about structural ambiguity and why it's hard for computers. After a brief introduction to structural ambiguity, users are challenged to complete a sentence in a way that tricks the computer into guessing an incorrect interpretation. Behind the scenes are two different NLP-based methods for classifying the user's input, one representative of classic rule-based approaches to disambiguation and the other representative of recent neural network approaches. Qualitative feedback from the system's use in online, classroom, and science museum settings indicates that it is engaging and successful in conveying the intended take home messages. A demo of Madly Ambiguous can be played at <http://madlyambiguous.osu.edu>.

VnCoreNLP: A Vietnamese Natural Language Processing Toolkit

Thanh Vu, Dat Quoc Nguyen, Dai Quoc Nguyen, Mark Dras, and Mark Johnson

We present an easy-to-use and fast toolkit, namely VnCoreNLP—a Java NLP annotation pipeline for Vietnamese. Our VnCoreNLP supports key natural language processing (NLP) tasks including word segmentation, part-of-speech (POS) tagging, named entity recognition (NER) and dependency parsing, and obtains state-of-the-art (SOTA) results for these tasks. We release VnCoreNLP to provide rich linguistic annotations to facilitate research work on Vietnamese NLP. Our VnCoreNLP is open-source and available at: <https://github.com/vncorenlp/VnCoreNLP>

Industry Keynote Address: Daniel Marcu

Building Innovative Startups, Products, and Services – Personal Insights

Sunday, June 3, 2018, 14:00–15:00

Chair: Jennifer Chu-Carroll

Empire Ballroom

Abstract: During the last 15 years, NLP&ML scientists have started to explore with increased persistence how to turn science into successful startups and how to incorporate cutting edge research into innovative products and services. In this talk, I will review lessons learned from my personal experience in this area.

Biography: Daniel is a Director of MT/NLP at Amazon. He is a 2014 fellow of the ACL for significant contributions to discourse parsing, summarization and machine translation and for kick starting the statistical machine translation industry.

Session 7 Overview – Sunday, June 3, 2018

	Track A <i>Text Mining 1</i>	Track B <i>Semantics 3</i>	Track C <i>Tagging, Chunking, Syntax and Parsing 2</i>	Track D <i>Industry: Panel</i>	Track P <i>Posters and Demos</i>
	Empire A	Empire B	Empire C	Empire D	Elite Hall B
15:30	Comparing Automatic and Human Evaluation of Local Explanations for Text Classification <i>D. Nguyen</i>	A Broad-Coverage Challenge Corpus for Sentence Understanding through Inference <i>A. Williams, N. Nangia, and S. Bowman</i>	Sentences with Gapping: Parsing and Reconstructing Elided Predicates <i>S. Schuster, J. Nivre, and C. D. Manning</i>		
15:48	Deep Temporal-Recurrent-Replicated-Sofmax for Topical Trends over Time <i>P. Gupta, S. Rajaram, B. Andrassy, and H. Schütze</i>	Filling Missing Paths: Modeling Co-occurrences of Word Pairs and Dependency Paths for Recognizing Lexical Semantic Relations <i>K. Washio and T. Kato</i>	A Structured Syntax-Semantics Interface for English-AMR Alignment <i>I. Szubert, A. Lopez, and N. Schneider</i>		
16:06	Lessons from the Bible on Modern Topics: Low-Resource Multilingual Topic Model Evaluation <i>S. Hao, J. Boyd-Graber, and M. J. Paul</i>	Specialising Word Vectors for Lexical Entailment <i>I. Vulić and N. Mrkšić</i>	End-to-end Graph-Based TAG Parsing with Neural Networks <i>J. Kasai, R. Frank, P. Xu, W. Merrill, and O. Rambow</i>	Industry Panel: Ethics in NLP	Machine Learning 4 / Machine Translation 2 / Phonology, Morphology and Word Segmentation 2
16:24	Explainable Prediction of Medical Codes from Clinical Text <i>J. Mullenbach, S. Wiegreffe, J. Duke, J. Sun, and J. Eisenstein</i>	Cross-Lingual Abstract Meaning Representation Parsing <i>M. Damonte and S. B. Cohen</i>	Colorless Green Recurrent Networks Dream Hierarchically <i>K. Gulordava, P. Bojanowski, E. Grave, T. Linzen, and M. Baroni</i>		
16:42	Event Time Extraction with a Decision Tree of Neural Classifiers (TACL) <i>N. Reimers, N. Dehghani, and I. Gurevych</i>	Mapping to Declarative Knowledge for Word Problem Solving (TACL) <i>S. Roy and D. Roth</i>	Knowledge Completion for Generics using Guided Tensor Factorization (TACL) <i>H. Sedghi and A. Sabharwal</i>		

Parallel Session 7

Oral: Text Mining 1

Empire A

Chair: Kai-Wei Chang

Comparing Automatic and Human Evaluation of Local Explanations for Text Classification *Dong Nguyen* 15:30–15:47

Text classification models are becoming increasingly complex and opaque, however for many applications it is essential that the models are interpretable. Recently, a variety of approaches have been proposed for generating local explanations. While robust evaluations are needed to drive further progress, so far it is unclear which evaluation approaches are suitable. This paper is a first step towards more robust evaluations of local explanations. We evaluate a variety of local explanation approaches using automatic measures based on word deletion. Furthermore, we show that an evaluation using a crowdsourcing experiment correlates moderately with these automatic measures and that a variety of other factors also impact the human judgements.

Deep Temporal-Recurrent-Replicated-Softmax for Topical Trends over Time

Pankaj Gupta, Subburam Rajaram, Bernt Andrassy, and Hinrich Schütze 15:48–16:05

Dynamic topic modeling facilitates the identification of topical trends over time in temporal collections of unstructured documents. We introduce a novel unsupervised neural dynamic topic model named as Recurrent Neural Network-Replicated Softmax Model (RNNRSM), where the discovered topics at each time influence the topic discovery in the subsequent time steps. We account for the temporal ordering of documents by explicitly modeling a joint distribution of latent topical dependencies over time, using distributional estimators with temporal recurrent connections. Applying RNN-RSM to 19 years of articles on NLP research, we demonstrate that compared to state-of-the-art topic models, RNNRSM shows better generalization, topic interpretation, evolution and trends. We also introduce a metric (named as SPAN) to quantify the capability of dynamic topic model to capture word evolution in topics over time.

Lessons from the Bible on Modern Topics: Low-Resource Multilingual Topic Model Evaluation

Shudong Hao, Jordan Boyd-Graber, and Michael J. Paul 16:06–16:23

Multilingual topic models enable document analysis across languages through coherent multilingual summaries of the data. However, there is no standard and effective metric to evaluate the quality of multilingual topics. We introduce a new intrinsic evaluation of multilingual topic models that correlates well with human judgments of multilingual topic coherence as well as performance in downstream applications. Importantly, we also study evaluation for low-resource languages. Because standard metrics fail to accurately measure topic quality when robust external resources are unavailable, we propose an adaptation model that improves the accuracy and reliability of these metrics in low-resource settings.

Explainable Prediction of Medical Codes from Clinical Text

James Mullenbach, Sarah Wiegrefe, Jon Duke, Jimeng Sun, and Jacob Eisenstein 16:24–16:41

Clinical notes are text documents that are created by clinicians for each patient encounter. They are typically accompanied by medical codes, which describe the diagnosis and treatment. Annotating these codes is labor intensive and error prone; furthermore, the connection between the codes and the text is not annotated, obscuring the reasons and details behind specific diagnoses and treatments. We present an attentional convolutional network that predicts medical codes from clinical text. Our method aggregates information across the document using a convolutional neural network, and uses an attention mechanism to select the most relevant segments for each of the thousands of possible codes. The method is accurate, achieving precision@8 of 0.71 and a Micro-F1 of 0.54, which are both better than the prior state of the art. Furthermore, through an interpretability evaluation by a physician, we show that the attention mechanism identifies meaningful explanations for each code assignment.

Event Time Extraction with a Decision Tree of Neural Classifiers (TACL)

Nils Reimers, Nazanin Dehghani, and Iryna Gurevych 16:42–17:00

Extracting the information from text when an event happened is challenging. Documents do not only report on current events, but also on past events as well as on future events. Often, the relevant time information for an event is scattered across the document. In this paper we present a novel method to automatically anchor events in time. To our knowledge it is the first approach that takes temporal information from the complete

document into account. We created a decision tree that applies neural network based classifiers at its nodes. It infers stepwise the final information at which date or at which time frame an event happened. We evaluate the approach on the TimeBank-EventTime Corpus (Reimers et al., 2016) achieving an accuracy of 42.0% compared to an inter-annotator agreement (IAA) of 56.7%. For events that span over a single day we observe an accuracy improvement of 33.1 points compared to the state-of-the-art CAEVO system (Chambers et al., 2014). Without re-training, we apply this model to the SemEval-2015 Task 4 on automatic timeline generation and achieve an improvement of 4.01 points F1-score compared to the state-of-the-art.

Oral: Semantics 3

Empire B

*Chair: Ido Dagan***A Broad-Coverage Challenge Corpus for Sentence Understanding through Inference***Adina Williams, Nikita Nangia, and Samuel Bowman*

15:30–15:47

This paper introduces the Multi-Genre Natural Language Inference (MultiNLI) corpus, a dataset designed for use in the development and evaluation of machine learning models for sentence understanding. At 433k examples, this resource is one of the largest corpora available for natural language inference (a.k.a. recognizing textual entailment), improving upon available resources in both its coverage and difficulty. MultiNLI accomplishes this by offering data from ten distinct genres of written and spoken English, making it possible to evaluate systems on nearly the full complexity of the language, while supplying an explicit setting for evaluating cross-genre domain adaptation. In addition, an evaluation using existing machine learning models designed for the Stanford NLI corpus shows that it represents a substantially more difficult task than does that corpus, despite the two showing similar levels of inter-annotator agreement.

Filling Missing Paths: Modeling Co-occurrences of Word Pairs and Dependency Paths for Recognizing Lexical Semantic Relations*Koki Washio and Tsuneki Kato*

15:48–16:05

Recognizing lexical semantic relations between word pairs is an important task for many applications of natural language processing. One of the mainstream approaches to this task is to exploit the lexico-syntactic paths connecting two target words, which reflect the semantic relations of word pairs. However, this method requires that the considered words co-occur in a sentence. This requirement is hardly satisfied because of Zipf's law, which states that most content words occur very rarely. In this paper, we propose novel methods with a neural model of $P(\text{path}_{\text{lw1}}, \text{w}_2)$ to solve this problem. Our proposed model of $P(\text{path}_{\text{lw1}}, \text{w}_2)$ can be learned in an unsupervised manner and can generalize the co-occurrences of word pairs and dependency paths. This model can be used to augment the path data of word pairs that do not co-occur in the corpus, and extract features capturing relational information from word pairs. Our experimental results demonstrate that our methods improve on previous neural approaches based on dependency paths and successfully solve the focused problem.

Specialising Word Vectors for Lexical Entailment*Ivan Vulić and Nikola Mrkšić*

16:06–16:23

We present LEAR (Lexical Entailment Attract-Repel), a novel post-processing method that transforms any input word vector space to emphasise the asymmetric relation of lexical entailment (LE), also known as the IS-A or hyponymy-hypernymy relation. By injecting external linguistic constraints (e.g., WordNet links) into the initial vector space, the LE specialisation procedure brings true hyponymy-hypernymy pairs closer together in the transformed Euclidean space. The proposed asymmetric distance measure adjusts the norms of word vectors to reflect the actual WordNet-style hierarchy of concepts. Simultaneously, a joint objective enforces semantic similarity using the symmetric cosine distance, yielding a vector space specialised for both lexical relations at once. LEAR specialisation achieves state-of-the-art performance in the tasks of hypernymy directionality, hypernymy detection, and graded lexical entailment, demonstrating the effectiveness and robustness of the proposed asymmetric specialisation model.

Cross-Lingual Abstract Meaning Representation Parsing*Marco Damonte and Shay B. Cohen*

16:24–16:41

Abstract Meaning Representation (AMR) research has mostly focused on English. We show that it is possible to use AMR annotations for English as a semantic representation for sentences written in other languages. We exploit an AMR parser for English and parallel corpora to learn AMR parsers for Italian, Spanish, German and Chinese. Qualitative analysis show that the new parsers overcome structural differences between the languages. We further propose a method to evaluate the parsers that does not require gold standard data in the target languages. This method highly correlates with the gold standard evaluation, obtaining a Pearson

correlation coefficient of 0.95.

Mapping to Declarative Knowledge for Word Problem Solving (TACL)

Subhra Roy and Dan Roth

16:42–17:00

Math word problems form a natural abstraction to a range of quantitative reasoning problems, such as understanding financial news, sports results, and casualties of war. Solving such problems requires the understanding of several mathematical concepts such as dimensional analysis, subset relationships, etc. In this paper, we develop declarative rules which govern the translation of natural language description of these concepts to math expressions. We then present a framework for incorporating such declarative knowledge into word problem solving. Our method learns to map arithmetic word problem text to math expressions, by learning to select the relevant declarative knowledge for each operation of the solution expression. This provides a way to handle multiple concepts in the same problem while, at the same time, support interpretability of the answer expression. Our method models the mapping to declarative knowledge as a latent variable, thus removing the need for expensive annotations. Experimental evaluation suggests that our domain knowledge based solver outperforms all other systems, and that it generalizes better in the realistic case where the training data it is exposed to is biased in a different way than the test data.

Oral: Tagging, Chunking, Syntax and Parsing 2

Empire C

Chair: Michael Collins

Sentences with Gapping: Parsing and Reconstructing Elided Predicates

Sebastian Schuster, Joakim Nivre, and Christopher D. Manning

15:30–15:47

Sentences with gapping, such as Paul likes coffee and Mary tea, lack an overt predicate to indicate the relation between two or more arguments. Surface syntax representations of such sentences are often produced poorly by parsers, and even if correct, not well suited to downstream natural language understanding tasks such as relation extraction that are typically designed to extract information from sentences with canonical clause structure. In this paper, we present two methods for parsing to Universal Dependencies graph representation that explicitly encodes the elided material with additional nodes and edges. We find that both methods can reconstruct elided material from dependency trees with high accuracy when the parser correctly predicts the existence of a gap. We further demonstrate that one of our methods can be applied to other languages based on a case study on Swedish.

A Structured Syntax-Semantics Interface for English-AMR Alignment

Ida Szubert, Adam Lopez, and Nathan Schneider

15:48–16:05

Abstract Meaning Representation (AMR) annotations are often assumed to closely mirror dependency syntax, but AMR explicitly does not require this, and the assumption has never been tested. To test it, we devise an expressive framework to align AMR graphs to dependency graphs, which we use to annotate 200 AMRs. Our annotation explains how 97% of AMR edges are evoked by words or syntax. Previously existing AMR alignment frameworks did not allow for mapping AMR onto syntax, and as a consequence they explained at most 23%. While we find that there are indeed many cases where AMR annotations closely mirror syntax, there are also pervasive differences. We use our annotations to test a baseline AMR-to-syntax aligner, finding that this task is more difficult than AMR-to-string alignment; and to pinpoint errors in an AMR parser. We make our data and code freely available for further research on AMR parsing and generation, and the relationship of AMR to syntax.

End-to-end Graph-Based TAG Parsing with Neural Networks

Jungo Kasai, Robert Frank, Pauli Xu, William Merrill, and Owen Rambow

16:06–16:23

We present a graph-based Tree Adjoining Grammar (TAG) parser that uses BiLSTMs, highway connections, and character-level CNNs. Our best end-to-end parser, which jointly performs supertagging, POS tagging, and parsing, outperforms the previously reported best results by more than 2.2 LAS and UAS points. The graph-based parsing architecture allows for global inference and rich feature representations for TAG parsing, alleviating the fundamental trade-off between transition-based and graph-based parsing systems. We also demonstrate that the proposed parser achieves state-of-the-art performance in the downstream tasks of Parsing Evaluation using Textual Entailments (PETE) and Unbounded Dependency Recovery. This provides further support for the claim that TAG is a viable formalism for problems that require rich structural analysis of sentences.

Colorless Green Recurrent Networks Dream Hierarchically

Kristina Gulordava, Piotr Bojanowski, Edouard Grave, Tal Linzen, and Marco Baroni 16:24–16:41

Recurrent neural networks (RNNs) achieved impressive results in a variety of linguistic processing tasks, suggesting that they can induce non-trivial properties of language. We investigate to what extent RNNs learn to track abstract hierarchical syntactic structure. We test whether RNNs trained with a generic language modeling objective in four languages (Italian, English, Hebrew, Russian) can predict long-distance number agreement in various constructions. We include in our evaluation nonsensical sentences where RNNs cannot rely on semantic or lexical cues (“The colorless green ideas I ate with the chair sleep furiously”), and, for Italian, we compare model performance to human intuitions. Our language-model-trained RNNs make reliable predictions about long-distance agreement, and do not lag much behind human performance. We thus bring support to the hypothesis that RNNs are not just shallow-pattern extractors, but they also acquire deeper grammatical competence.

Knowledge Completion for Generics using Guided Tensor Factorization (TACL)

Hanie Sedghi and Ashish Sabharwal

16:42–17:00

Given a knowledge base or KB containing (noisy) facts about common nouns or generics, such as “all trees produce oxygen” or “some animals live in forests”, we consider the problem of inferring additional such facts at a precision similar to that of the starting KB. Such KBs capture general knowledge about the world, and are crucial for various applications such as question answering. Different from commonly studied named entity KBs such as Freebase, generics KBs involve quantification, have more complex underlying regularities, tend to be more incomplete, and violate the commonly used locally closed world assumption (LCWA). We show that existing KB completion methods struggle with this new task, and present the first approach that is successful. Our results demonstrate that external information, such as relation schemas and entity taxonomies, if used appropriately, can be a surprisingly powerful tool in this setting. First, our simple yet effective knowledge guided tensor factorization approach achieves state-of-the-art results on two generics KBs (80% precise) for science, doubling their size at 74%-86% precision. Second, our novel taxonomy guided, submodular, active learning method for collecting annotations about rare entities (e.g., oriole, a bird) is 6x more effective at inferring further new facts about them than multiple active learning baselines.

Industry Panel: Ethics in NLP

Empire D

Chair: Dirk Hovy

Posters and Demos: Session 7

Posters: Machine Learning 4

Time: 15:30–17:00

Location: Elite Hall B

Chair: *Mona Diab*

Pivot Based Language Modeling for Improved Neural Domain Adaptation

Yftah Ziser and Roi Reichart

Representation learning with pivot-based methods and with Neural Networks (NNs) have lead to significant progress in domain adaptation for Natural Language Processing. However, most previous work that follows these approaches does not explicitly exploit the structure of the input text, and its output is most often a single representation vector for the entire text. In this paper we present the Pivot Based Language Model (PBLM), a representation learning model that marries together pivot-based and NN modeling in a structure aware manner. Particularly, our model processes the information in the text with a sequential NN (LSTM) and its output consists of a representation vector for every input word. Unlike most previous representation learning models in domain adaptation, PBLM can naturally feed structure aware text classifiers such as LSTM and CNN. We experiment with the task of cross-domain sentiment classification on 20 domain pairs and show substantial improvements over strong baselines.

Multinomial Adversarial Networks for Multi-Domain Text Classification

Xilun Chen and Claire Cardie

Many text classification tasks are known to be highly domain-dependent. Unfortunately, the availability of training data can vary drastically across domains. Worse still, for some domains there may not be any annotated data at all. In this work, we propose a multinomial adversarial network (MAN) to tackle this real-world problem of multi-domain text classification (MDTC) in which labeled data may exist for multiple domains, but in insufficient amounts to train effective classifiers for one or more of the domains. We provide theoretical justifications for the MAN framework, proving that different instances of MANs are essentially minimizers of various f-divergence metrics (Ali and Silvey, 1966) among multiple probability distributions. MANs are thus a theoretically sound generalization of traditional adversarial networks that discriminate over two distributions. More specifically, for the MDTC task, MAN learns features that are invariant across multiple domains by resorting to its ability to reduce the divergence among the feature distributions of each domain. We present experimental results showing that MANs significantly outperform the prior art on the MDTC task. We also show that MANs achieve state-of-the-art performance for domains with no labeled data.

Early Text Classification Using Multi-Resolution Concept Representations

Adrián Pastor López Monroy, Fabio A. González, Hugo Jair Escalante, Manuel Montes, and Thamar Solorio

The intensive use of e-communications in everyday life has given rise to new threats and risks. When the vulnerable asset is the user, detecting these potential attacks before they cause serious damages is extremely important. This paper proposes a novel document representation to improve the early detection of risks in social media sources. The goal is to effectively identify the potential risk using as few text as possible and with as much anticipation as possible. Accordingly, we devise a Multi-Resolution Representation (MuRR), which allows us to generate multiple “views” of the analyzed text. These views capture different semantic meanings for words and documents at different levels of detail, which is very useful in early scenarios to model the variable amounts of evidence. Intuitively, the representation captures better the content of short documents (very early stages) in low resolutions, whereas large documents (medium/large stages) are better modeled with higher resolutions. We evaluate the proposed ideas in two different tasks where anticipation is critical: sexual predator detection and depression detection. The experimental evaluation for these early tasks revealed that the proposed approach outperforms previous methodologies by a considerable margin.

The Context-dependent Additive Recurrent Neural Net

Quan Hung Tran, Tuan Lai, Ingrid Zukerman, Gholamreza Haffari, Trung Bui, and Hung Bui

Contextual sequence mapping is one of the fundamental problems in Natural Language Processing (NLP). Here, instead of relying solely on the information presented in the text, the learning agents have access to a strong external signal given to assist the learning process. In this paper, we propose a novel family of Recurrent Neural Network unit: the Context-dependent Additive Recurrent Neural Network (CARNN) that is designed specifically to address this type of problem. The experimental results on public datasets in the dialog problem

(Babi dialog Task 6 and Frame), contextual language model (Switchboard and Penn Tree Bank) and question answering (Trec QA) show that our novel CARNN-based architectures outperform previous methods.

Diverse Few-Shot Text Classification with Multiple Metrics

Mo Yu, Xiaoxiao Guo, Jinfeng Yi, Shiyu Chang, Saloni Potdar, Yu Cheng, Gerald Tesauro, Haoyu Wang, and Bowen Zhou

We study few-shot learning in natural language domains. Compared to many existing works that apply either metric-based or optimization-based meta-learning to image domain with low inter-task variance, we consider a more realistic setting, where tasks are diverse. However, it imposes tremendous difficulties to existing state-of-the-art metric-based algorithms since a single metric is insufficient to capture complex task variations in natural language domain. To alleviate the problem, we propose an adaptive metric learning approach that automatically determines the best weighted combination from a set of metrics obtained from meta-training tasks for a newly seen few-shot task. Extensive quantitative evaluations on real-world sentiment analysis and dialog intent classification datasets demonstrate that the proposed method performs favorably against state-of-the-art few shot learning algorithms in terms of predictive accuracy. We make our code and data available for further study.

Reinforced Co-Training

Jiawei Wu, Lei Li, and William Yang Wang

Co-training is a popular semi-supervised learning framework to utilize a large amount of unlabeled data in addition to a small labeled set. Co-training methods exploit predicted labels on the unlabeled data and select samples based on prediction confidence to augment the training. However, the selection of samples in existing co-training methods is based on a predetermined policy, which ignores the sampling bias between the unlabeled and the labeled subsets, and fails to explore the data space. In this paper, we propose a novel method, Reinforced Co-Training, to select high-quality unlabeled samples to better co-train on. More specifically, our approach uses Q-learning to learn a data selection policy with a small labeled dataset, and then exploits this policy to train the co-training classifiers automatically. Experimental results on clickbait detection and generic text classification tasks demonstrate that our proposed method can obtain more accurate text classification results.

Contextual Augmentation: Data Augmentation by Words with Paradigmatic Relations

Sosuke Kobayashi

We propose a novel data augmentation for labeled sentences called contextual augmentation. We assume an invariance that sentences are natural even if the words in the sentences are replaced with other words with paradigmatic relations. We stochastically replace words with other words that are predicted by a bi-directional language model at the word positions. Words predicted according to a context are numerous but appropriate for the augmentation of the original words. Furthermore, we retrofit a language model with a label-conditional architecture, which allows the model to augment sentences without breaking the label-compatibility. Through the experiments for six various different text classification tasks, we demonstrate that the proposed method improves classifiers based on the convolutional or recurrent neural networks.

Cross-Lingual Learning-to-Rank with Shared Representations

Shota Sasaki, Shuo Sun, Shigehiko Schamoni, Kevin Duh, and Kentaro Inui

Cross-lingual information retrieval (CLIR) is a document retrieval task where the documents are written in a language different from that of the user's query. This is a challenging problem for data-driven approaches due to the general lack of labeled training data. We introduce a large-scale dataset derived from Wikipedia to support CLIR research in 25 languages. Further, we present a simple yet effective neural learning-to-rank model that shares representations across languages and reduces the data requirement. This model can exploit training data in, for example, Japanese-English CLIR to improve the results of Swahili-English CLIR.

Text Segmentation as a Supervised Learning Task

Omri Koshorek, Adir Cohen, Noam Mor, Michael Rotman, and Jonathan Berant

Text segmentation, the task of dividing a document into contiguous segments based on its semantic structure, is a longstanding challenge in language understanding. Previous work on text segmentation focused on unsupervised methods such as clustering or graph search, due to the paucity in labeled data. In this work, we formulate text segmentation as a supervised learning problem, and present a large new dataset for text segmentation that is automatically extracted and labeled from Wikipedia. Moreover, we develop a segmentation model based on this dataset and show that it generalizes well to unseen natural text.

Self-Attention with Relative Position Representations*Peter Shaw, Jakob Uszkoreit, and Ashish Vaswani*

Relying entirely on an attention mechanism, the Transformer introduced by Vaswani et al. (2017) achieves state-of-the-art results for machine translation. In contrast to recurrent and convolutional neural networks, it does not explicitly model relative or absolute position information in its structure. Instead, it requires adding representations of absolute positions to its inputs. In this work we present an alternative approach, extending the self-attention mechanism to efficiently consider representations of the relative positions, or distances between sequence elements. On the WMT 2014 English-to-German and English-to-French translation tasks, this approach yields improvements of 1.3 BLEU and 0.3 BLEU over absolute position representations, respectively. Notably, we observe that combining relative and absolute position representations yields no further improvement in translation quality. We describe an efficient implementation of our method and cast it as an instance of relation-aware self-attention mechanisms that can generalize to arbitrary graph-labeled inputs.

What's in a Domain? Learning Domain-Robust Text Representations using Adversarial Training*Yitong Li, Timothy Baldwin, and Trevor Cohn*

Most real world language problems require learning from heterogenous corpora, raising the problem of learning robust models which generalise well to both similar (*in domain*) and dissimilar (*out of domain*) instances to those seen in training. This requires learning an underlying task, while not learning irrelevant signals and biases specific to individual domains. We propose a novel method to optimise both in- and out-of-domain accuracy based on joint learning of a structured neural model with domain-specific and domain-general components, coupled with adversarial training for domain. Evaluating on multi-domain language identification and multi-domain sentiment analysis, we show substantial improvements over standard domain adaptation techniques, and domain-adversarial training.

Tensor Product Generation Networks for Deep NLP Modeling*Qiuyuan Huang, Paul Smolensky, Xiaodong He, Li Deng, and Dapeng Wu*

We present a new approach to the design of deep networks for natural language processing (NLP), based on the general technique of Tensor Product Representations (TPRs) for encoding and processing symbol structures in distributed neural networks. A network architecture — the Tensor Product Generation Network (TPGN) — is proposed which is capable in principle of carrying out TPR computation, but which uses unconstrained deep learning to design its internal representations. Instantiated in a model for image-caption generation, TPGN outperforms LSTM baselines when evaluated on the COCO dataset. The TPR-capable structure enables interpretation of internal representations and operations, which prove to contain considerable grammatical content. Our caption-generation model can be interpreted as generating sequences of grammatical categories and retrieving words by their categories from a plan encoded as a distributed representation.

ListOps: A Diagnostic Dataset for Latent Tree Learning (SRW)*Nikita Nangia and Samuel Bowman*

Latent tree learning models learn to parse a sentence without syntactic supervision, and use that parse to build the sentence representation. Existing work on such models has shown that, while they perform well on tasks like sentence classification, they do not learn grammars that conform to any plausible semantic or syntactic formalism (Williams et al., 2018a). Studying the parsing ability of such models in natural language can be challenging due to the inherent complexities of natural language, like having several valid parses for a single sentence. In this paper we introduce ListOps, a toy dataset created to study the parsing ability of latent tree models. ListOps sequences are in the style of prefix arithmetic. The dataset is designed to have a single correct parsing strategy that a system needs to learn to succeed at the task. We show that the current leading latent tree models are unable to learn to parse and succeed at ListOps. These models achieve accuracies worse than purely sequential RNNs.

Neural Lattice Language Models (TACL)*Jacob Buckman and Graham Neubig*

In this work, we propose a new language modeling paradigm that has the ability to perform both prediction and moderation of information flow at multiple granularities: neural lattice language models. These models construct a lattice of possible paths through a sentence and marginalize across this lattice to calculate sequence probabilities or optimize parameters. This approach allows us to seamlessly incorporate linguistic intuitions – including polysemy and existence of multi-word lexical items – into our language model. Experiments on multiple language modeling tasks show that over English neural lattice language models that utilize polysemous embeddings are able to improve perplexity by 9.95% relative to a word-level baseline, and that a Chinese

model that handles multi-character tokens is able to improve perplexity by 20.94% relative to a character-level baseline.

Practical Application of Domain Dependent Confidence Measurement for Spoken Language Understanding Systems (INDUSTRY)

Mahnoush Mehrabani, David Thomson, and Benjamin Stern

Spoken Language Understanding (SLU), which extracts semantic information from speech, is not flawless, especially in practical applications. The reliability of the output of an SLU system can be evaluated using a semantic confidence measure. Confidence measures are a solution to improve the quality of spoken dialogue systems, by rejecting low-confidence SLU results. In this study we discuss real-world applications of confidence scoring in a customer service scenario. We build confidence models for three major types of dialogue states that are considered as different domains: how may I help you, number capture, and confirmation. Practical challenges to train domain-dependent confidence models, including data limitations, are discussed, and it is shown that feature engineering plays an important role to improve performance. We explore a wide variety of predictor features based on speech recognition, intent classification, and high-level domain knowledge, and find the combined feature set with the best rejection performance for each application.

Posters: Machine Translation 2

Time: 15:30–17:00

Location: Elite Hall B

Chair: *Mona Diab*

Evaluating Discourse Phenomena in Neural Machine Translation

Rachel Bawden, Rico Sennrich, Alexandra Birch, and Barry Haddow

For machine translation to tackle discourse phenomena, models must have access to extra-sentential linguistic context. There has been recent interest in modelling context in neural machine translation (NMT), but models have been principally evaluated with standard automatic metrics, poorly adapted to evaluating discourse phenomena. In this article, we present hand-crafted, discourse test sets, designed to test the models' ability to exploit previous source and target sentences. We investigate the performance of recently proposed multi-encoder NMT models trained on subtitles for English to French. We also explore a novel way of exploiting context from the previous sentence. Despite gains using BLEU, multi-encoder models give limited improvement in the handling of discourse phenomena: 50% accuracy on our coreference test set and 53.5% for coherence/cohesion (compared to a non-contextual baseline of 50%). A simple strategy of decoding the concatenation of the previous and current sentence leads to good performance, and our novel strategy of multi-encoding and decoding of two sentences leads to the best performance (72.5% for coreference and 57% for coherence/cohesion), highlighting the importance of target-side context.

Improving Neural Machine Translation with Conditional Sequence Generative Adversarial Nets

Zhen Yang, Wei Chen, Feng Wang, and Bo Xu

This paper proposes an approach for applying GANs to NMT. We build a conditional sequence generative adversarial net which comprises of two adversarial sub models, a generator and a discriminator. The generator aims to generate sentences which are hard to be discriminated from human-translated sentences (i.e., the golden target sentences); And the discriminator makes efforts to discriminate the machine-generated sentences from human-translated ones. The two sub models play a mini-max game and achieve the win-win situation when they reach a Nash Equilibrium. Additionally, the static sentence-level BLEU is utilized as the reinforced objective for the generator, which biases the generation towards high BLEU points. During training, both the dynamic discriminator and the static BLEU objective are employed to evaluate the generated sentences and feedback the evaluations to guide the learning of the generator. Experimental results show that the proposed model consistently outperforms the traditional RNNSearch and the newly emerged state-of-the-art Transformer on English-German and Chinese-English translation tasks.

Self-Attentive Residual Decoder for Neural Machine Translation

Lesly Miculicich Werlen, Nikolaos Pappas, Dhananjay Ram, and Andrei Popescu-Belis

Neural sequence-to-sequence networks with attention have achieved remarkable performance for machine translation. One of the reasons for their effectiveness is their ability to capture relevant source-side contextual information at each time-step prediction through an attention mechanism. However, the target-side context is solely based on the sequence model which, in practice, is prone to a recency bias and lacks the ability to capture effectively non-sequential dependencies among words. To address this limitation, we propose a target-side-

attentive residual recurrent network for decoding, where attention over previous words contributes directly to the prediction of the next word. The residual learning facilitates the flow of information from the distant past and is able to emphasize any of the previously translated words, hence it gains access to a wider context. The proposed model outperforms a neural MT baseline as well as a memory and self-attention network on three language pairs. The analysis of the attention learned by the decoder confirms that it emphasizes a wider context, and that it captures syntactic-like structures.

Handling Homographs in Neural Machine Translation

Frederick Liu, Han Lu, and Graham Neubig

Homographs, words with different meanings but the same surface form, have long caused difficulty for machine translation systems, as it is difficult to select the correct translation based on the context. However, with the advent of neural machine translation (NMT) systems, which can theoretically take into account global sentential context, one may hypothesize that this problem has been alleviated. In this paper, we first provide empirical evidence that existing NMT systems in fact still have significant problems in properly translating ambiguous words. We then proceed to describe methods, inspired by the word sense disambiguation literature, that model the context of the input word with context-aware word embeddings that help to differentiate the word sense before feeding it into the encoder. Experiments on three language pairs demonstrate that such models improve the performance of NMT systems both in terms of BLEU score and in the accuracy of translating homographs.

Dense Information Flow for Neural Machine Translation

Yanyao Shen, Xu Tan, Di He, Tao Qin, and Tie-Yan Liu

Recently, neural machine translation has achieved remarkable progress by introducing well-designed deep neural networks into its encoder-decoder framework. From the optimization perspective, residual connections are adopted to improve learning performance for both encoder and decoder in most of these deep architectures, and advanced attention connections are applied as well. Inspired by the success of the DenseNet model in computer vision problems, in this paper, we propose a densely connected NMT architecture (DenseNMT) that is able to train more efficiently for NMT. The proposed DenseNMT not only allows dense connection in creating new features for both encoder and decoder, but also uses the dense attention structure to improve attention quality. Our experiments on multiple datasets show that DenseNMT structure is more competitive and efficient.

Target Foresight Based Attention for Neural Machine Translation

Xintong Li, Lemaio Liu, Zhaopeng Tu, Shuming Shi, and Max Meng

In neural machine translation, an attention model is used to identify the aligned source words for a target word (target foresight word) in order to select translation context, but it does not make use of any information of this target foresight word at all. Previous work proposed an approach to improve the attention model by explicitly accessing this target foresight word and demonstrated the substantial gains in alignment task. However, this approach is useless in machine translation task on which the target foresight word is unavailable. In this paper, we propose a new attention model enhanced by the implicit information of target foresight word oriented to both alignment and translation tasks. Empirical experiments on Chinese-to-English and Japanese-to-English datasets show that the proposed attention model delivers significant improvements in terms of both alignment error rate and BLEU.

Fast Lexically Constrained Decoding with Dynamic Beam Allocation for Neural Machine Translation

Matt Post and David Vilar

The end-to-end nature of neural machine translation (NMT) removes many ways of manually guiding the translation process that were available in older paradigms. Recent work, however, has introduced a new capability: lexically constrained or guided decoding, a modification to beam search that forces the inclusion of pre-specified words and phrases in the output. However, while theoretically sound, existing approaches have computational complexities that are either linear (Hokamp and Liu, 2017) or exponential (Anderson et al., 2017) in the number of constraints. We present an algorithm for lexically constrained decoding with a complexity of $O(1)$ in the number of constraints. We demonstrate the algorithm’s remarkable ability to properly place these constraints, and use it to explore the shaky relationship between model and BLEU scores. Our implementation is available as part of Sockeye.

Guiding Neural Machine Translation with Retrieved Translation Pieces

Jingyi Zhang, Masao Utiyama, Eiichiro Sumita, Graham Neubig, and Satoshi Nakamura

One of the difficulties of neural machine translation (NMT) is the recall and appropriate translation of low-

frequency words or phrases. In this paper, we propose a simple, fast, and effective method for recalling previously seen translation examples and incorporating them into the NMT decoding process. Specifically, for an input sentence, we use a search engine to retrieve sentence pairs whose source sides are similar with the input sentence, and then collect n-grams that are both in the retrieved target sentences and aligned with words that match in the source sentences, which we call “translation pieces”. We compute pseudo-probabilities for each retrieved sentence based on similarities between the input sentence and the retrieved source sentences, and use these to weight the retrieved translation pieces. Finally, an existing NMT model is used to translate the input sentence, with an additional bonus given to outputs that contain the collected translation pieces. We show our method improves NMT translation results up to 6 BLEU points on three narrow domain translation tasks where repetitiveness of the target sentences is particularly salient. It also causes little increase in the translation time, and compares favorably to another alternative retrieval-based method with respect to accuracy, speed, and simplicity of implementation.

Neural Machine Translation for Bilingually Scarce Scenarios: a Deep Multi-Task Learning Approach*Poorya Zaremodi and Gholamreza Haffari*

Neural machine translation requires large amount of parallel training text to learn a reasonable quality translation model. This is particularly inconvenient for language pairs for which enough parallel text is not available. In this paper, we use monolingual linguistic resources in the source side to address this challenging problem based on a multi-task learning approach. More specifically, we scaffold the machine translation task on auxiliary tasks including semantic parsing, syntactic parsing, and named-entity recognition. This effectively injects semantic and/or syntactic knowledge into the translation model, which would otherwise require a large amount of training bitext to learn from. We empirically analyze and show the effectiveness of our multitask learning approach on three translation tasks: English-to-French, English-to-Farsi, and English-to-Vietnamese.

Combining Character and Word Information in Neural Machine Translation Using a Multi-Level Attention*Huadong Chen, Shujian Huang, David Chiang, Xin-Yu Dai, and Jiajun Chen*

Natural language sentences, being hierarchical, can be represented at different levels of granularity, like words, subwords, or characters. But most neural machine translation systems require the sentence to be represented as a sequence at a single level of granularity. It can be difficult to determine which granularity is better for a particular translation task. In this paper, we improve the model by incorporating multiple levels of granularity. Specifically, we propose (1) an encoder with character attention which augments the (sub)word-level representation with character-level information; (2) a decoder with multiple attentions that enable the representations from different levels of granularity to control the translation cooperatively. Experiments on three translation tasks demonstrate that our proposed models outperform the standard word-based model, the subword-based model, and a strong character-based model.

Using Word Vectors to Improve Word Alignments for Low Resource Machine Translation*Nima Pourdamghani, Marjan Ghazvininejad, and Kevin Knight*

We present a method for improving word alignments using word similarities. This method is based on encouraging common alignment links between semantically similar words. We use word vectors trained on monolingual data to estimate similarity. Our experiments on translating fifteen languages into English show consistent BLEU score improvements across the languages.

Neural Machine Translation Decoding with Terminology Constraints*Eva Haster, Adrià de Gispert, Gonzalo Iglesias, and Bill Byrne*

Despite the impressive quality improvements yielded by neural machine translation (NMT) systems, controlling their translation output to adhere to user-provided terminology constraints remains an open problem. We describe our approach to constrained neural decoding based on finite-state machines and multi-stack decoding which supports target-side constraints as well as constraints with corresponding aligned input text spans. We demonstrate the performance of our framework on multiple translation tasks and motivate the need for constrained decoding with attentions as a means of reducing misplacement and duplication when translating user constraints.

When and Why Are Pre-Trained Word Embeddings Useful for Neural Machine Translation?*Ye Qi, Devendra Sachan, Matthieu Felix, Sarguna Padmanabhan, and Graham Neubig*

The performance of Neural Machine Translation (NMT) systems often suffers in low-resource scenarios where sufficiently large-scale parallel corpora cannot be obtained. Pre-trained word embeddings have proven to be

invaluable for improving performance in natural language analysis tasks, which often suffer from paucity of data. However, their utility for NMT has not been extensively explored. In this work, we perform five sets of experiments that analyze when we can expect pre-trained word embeddings to help in NMT tasks. We show that such embeddings can be surprisingly effective in some cases – providing gains of up to 20 BLEU points in the most favorable setting.

Automated Paraphrase Lattice Creation for HyTER Machine Translation Evaluation

Marianna Apidianaki, Guillaume Wisniewski, Anne Cocos, and Chris Callison-Burch

We propose a variant of a well-known machine translation (MT) evaluation metric, HyTER (Dreyer and Marcu, 2012), which exploits reference translations enriched with meaning equivalent expressions. The original HyTER metric relied on hand-crafted paraphrase networks which restricted its applicability to new data. We test, for the first time, HyTER with automatically built paraphrase lattices. We show that although the metric obtains good results on small and carefully curated data with both manually and automatically selected substitutes, it achieves medium performance on much larger and noisier datasets, demonstrating the limits of the metric for tuning and evaluation of current MT systems.

Learning Hidden Unit Contribution for Adapting Neural Machine Translation Models

David Vilar

In this paper we explore the use of Learning Hidden Unit Contribution for the task of neural machine translation. The method was initially proposed in the context of speech recognition for adapting a general system to the specific acoustic characteristics of each speaker. Similar in spirit, in a machine translation framework we want to adapt a general system to a specific domain. We show that the proposed method achieves improvements of up to 2.6 BLEU points over a general system, and up to 6 BLEU points if the initial system has only been trained on out-of-domain data, a situation which may easily happen in practice. The good performance together with its short training time and small memory footprint make it a very attractive solution for domain adaptation.

Exploiting Semantics in Neural Machine Translation with Graph Convolutional Networks

Diego Marcheggiani, Joost Bastings, and Ivan Titov

Semantic representations have long been argued as potentially useful for enforcing meaning preservation and improving generalization performance of machine translation methods. In this work, we are the first to incorporate information about predicate–argument structure of source sentences (namely, semantic-role representations) into neural machine translation. We use Graph Convolutional Networks (GCNs) to inject a semantic bias into sentence encoders and achieve improvements in BLEU scores over the linguistic-agnostic and syntax-aware versions on the English–German language pair.

On the Evaluation of Semantic Phenomena in Neural Machine Translation Using Natural Language Inference

Adam Poliak, Yonatan Belinkov, James Glass, and Benjamin Van Durme

We propose a process for investigating the extent to which sentence representations arising from neural machine translation (NMT) systems encode distinct semantic phenomena. We use these representations as features to train a natural language inference (NLI) classifier based on datasets recast from existing semantic annotations. In applying this process to a representative NMT system, we find its encoder appears most suited to supporting inferences at the syntax–semantics interface, as compared to anaphora resolution requiring world knowledge. We conclude with a discussion on the merits and potential deficiencies of the existing process, and how it may be improved and extended as a broader framework for evaluating semantic coverage.

Incremental Decoding and Training Methods for Simultaneous Translation in Neural Machine Translation

Fahim Dalvi, Nadir Durrani, Hassan Sajjad, and Stephan Vogel

We address the problem of simultaneous translation by modifying the Neural MT decoder to operate with dynamically built encoder and attention. We propose a tunable agent which decides the best segmentation strategy for a user-defined BLEU loss and Average Proportion (AP) constraint. Our agent outperforms previously proposed Wait-if-diff and Wait-if-worse agents (Cho and Esipova, 2016) on BLEU with a lower latency. Secondly we proposed data-driven changes to Neural MT training to better match the incremental decoding framework.

Session 7: Posters and Demos: Phonology, Morphology and Word Segmentation 2

Japanese Predicate Conjugation for Neural Machine Translation (SRW) *Michiki Kuroawa, Yukio Matsumura, Hayahide Yamagishi, and Mamoru Komachi*

Neural machine translation (NMT) has a drawback in that can generate only high-frequency words owing to the computational costs of the softmax function in the output layer. In Japanese-English NMT, Japanese predicate conjugation causes an increase in vocabulary size. For example, one verb can have as many as 19 surface varieties. In this research, we focus on predicate conjugation for compressing the vocabulary size in Japanese. The vocabulary list is filled with the various forms of verbs. We propose methods using predicate conjugation information without discarding linguistic information. The proposed methods can generate low-frequency words and deal with unknown words. Two methods were considered to introduce conjugation information: the first considers it as a token (conjugation token) and the second considers it as an embedded vector (conjugation feature). The results using these methods demonstrate that the vocabulary size can be compressed by approximately 86.1% (Tanaka corpus) and the NMT models can output the words not in the training data set. Furthermore, BLEU scores improved by 0.91 points in Japanese-to-English translation, and 0.32 points in English-to-Japanese translation with ASPEC.

Metric for Automatic Machine Translation Evaluation based on Universal Sentence Representations (SRW)

Hiroki Shimanaka, Tomoyuki Kajiwara, and Mamoru Komachi

Sentence representations can capture a wide range of information that cannot be captured by local features based on character or word N-grams. This paper examines the usefulness of universal sentence representations for evaluating the quality of machine translation. Although it is difficult to train sentence representations using small-scale translation datasets with manual evaluation, sentence representations trained from large-scale data in other tasks can improve the automatic evaluation of machine translation. Experimental results of the WMT-2016 dataset show that the proposed method achieves state-of-the-art performance with sentence representation features only.

Neural Machine Translation for Low Resource Languages using Bilingual Lexicon Induced from Comparable Corpora (SRW)

Sree Harsha Ramesh and Krishna Prasad Sankaranarayanan

Resources for the non-English languages are scarce and this paper addresses this problem in the context of machine translation, by automatically extracting parallel sentence pairs from the multilingual articles available on the Internet. In this paper, we have used an end-to-end Siamese bidirectional recurrent neural network to generate parallel sentences from comparable multilingual articles in Wikipedia. Subsequently, we have showed that using the harvested dataset improved BLEU scores on both NMT and phrase-based SMT systems for the low-resource language pairs: English—Hindi and English—Tamil, when compared to training exclusively on the limited bilingual corpora collected for these language pairs.

Posters: Phonology, Morphology and Word Segmentation 2

Time: 15:30–17:00

Location: Elite Hall B

Chair: Mona Diab

Reusing Weights in Subword-Aware Neural Language Models

Zhenisbek Assylbekov and Rustem Takhanov

We propose several ways of reusing subword embeddings and other weights in subword-aware neural language models. The proposed techniques do not benefit a competitive character-aware model, but some of them improve the performance of syllable- and morpheme-aware models while showing significant reductions in model sizes. We discover a simple hands-on principle: in a multi-layer input embedding model, layers should be tied consecutively bottom-up if reused at output. Our best morpheme-aware model with properly reused weights beats the competitive word-level model by a large margin across multiple languages and has 20%-87% fewer parameters.

Context Sensitive Neural Lemmatization with Lematus

Toms Bergmanis and Sharon Goldwater

The main motivation for developing contextsensitive lemmatizers is to improve performance on unseen and ambiguous words. Yet previous systems have not carefully evaluated whether the use of context actually helps in these cases. We introduce Lematus, a lemmatizer based on a standard encoder-decoder architecture, which incorporates character-level sentence context. We evaluate its lemmatization accuracy across 20 languages in both a full data setting and a lower-resource setting with 10k training examples in each language. In both

settings, we show that including context significantly improves results against a context-free version of the model. Context helps more for ambiguous words than for unseen words, though the latter has a greater effect on overall performance differences between languages. We also compare to three previous context-sensitive lemmatization systems, which all use pre-extracted edit trees as well as hand-selected features and/or additional sources of information such as tagged training data. Without using any of these, our context-sensitive model outperforms the best competitor system (Lemming) in the fulldata setting, and performs on par in the lowerresource setting.

Simple Models for Word Formation in Slang

Vivek Kulkarni and William Yang Wang

We propose the first generative models for three types of extra-grammatical word formation phenomena abounding in slang: Blends, Clippings, and Reduplicatives. Adopting a data-driven approach coupled with linguistic knowledge, we propose simple models with state of the art performance on human annotated gold standard datasets. Overall, our models reveal insights into the generative processes of word formation in slang – insights which are increasingly relevant in the context of the rising prevalence of slang and non-standard varieties on the Internet

Using Morphological Knowledge in Open-Vocabulary Neural Language Models

Austin Matthews, Graham Neubig, and Chris Dyer

Languages with productive morphology pose problems for language models that generate words from a fixed vocabulary. Although character-based models allow any possible word type to be generated, they are linguistically naïve: they must discover that words exist and are delimited by spaces—basic linguistic facts that are built in to the structure of word-based models. We introduce an open-vocabulary language model that incorporates more sophisticated linguistic knowledge by predicting words using a mixture of three generative processes: (1)–by generating words as a sequence of characters, (2)–by directly generating full word forms, and (3)–by generating words as a sequence of morphemes that are combined using a hand-written morphological analyzer. Experiments on Finnish, Turkish, and Russian show that our model outperforms character sequence models and other strong baselines on intrinsic and extrinsic measures. Furthermore, we show that our model learns to exploit morphological knowledge encoded in the analyzer, and, as a byproduct, it can perform effective unsupervised morphological disambiguation.

Unsupervised Disambiguation of Syncretism in Inflected Lexicons

Ryan Cotterell, Christo Kirov, Sebastian J. Mielke, and Jason Eisner

Lexical ambiguity makes it difficult to compute useful statistics of a corpus. A given word form might represent any of several morphological feature bundles. One can, however, use unsupervised learning (as in EM) to fit a model that probabilistically disambiguates word forms. We present such an approach, which employs a neural network to smoothly model a prior distribution over feature bundles (even rare ones). Although this basic model does not consider a token’s context, that very property allows it to operate on a simple list of unigram type counts, partitioning each count among different analyses of that unigram. We discuss evaluation metrics for this novel task and report results on 5 languages.

Are All Languages Equally Hard to Language-Model?

Ryan Cotterell, Sebastian J. Mielke, Jason Eisner, and Brian Roark

For general modeling methods applied to diverse languages, a natural question is: how well should we expect our models to work on languages with differing typological profiles? In this work, we develop an evaluation framework for fair cross-linguistic comparison of language models, using translated text so that all models are asked to predict approximately the same information. We then conduct a study on 21 languages, demonstrating that in some languages, the textual expression of the information is harder to predict with both n-gram and LSTM language models. We show complex inflectional morphology to be a cause of performance differences among languages.

Modeling Noisiness to Recognize Named Entities using Multitask Neural Networks on Social Media

Gustavo Aguilar, Adrian Pastor López Monroy, Fabio A. González, and Thamar Solorio

Recognizing named entities in a document is a key task in many NLP applications. Although current state-of-the-art approaches to this task reach a high performance on clean text (e.g. newswire genres), those algorithms dramatically degrade when they are moved to noisy environments such as social media domains. We present two systems that address the challenges of processing social media data using character-level phonetics and phonology, word embeddings, and Part-of-Speech tags as features. The first model is a multitask end-to-end

Bidirectional Long Short-Term Memory (BLSTM)-Conditional Random Field (CRF) network whose output layer contains two CRF classifiers. The second model uses a multitask BLSTM network as feature extractor that transfers the learning to a CRF classifier for the final prediction. Our systems outperform the current F1 scores of the state of the art on the Workshop on Noisy User-generated Text 2017 dataset by 2.45% and 3.69%, establishing a more suitable approach for social media environments.

The Computational Complexity of Distinctive Feature Minimization in Phonology*Hubie Chen and Mans Hulden*

We analyze the complexity of the problem of determining whether a set of phonemes forms a natural class and, if so, that of finding the minimal feature specification for the class. A standard assumption in phonology is that finding a minimal feature specification is an automatic part of acquisition and generalization. We find that the natural class decision problem is tractable (i.e. is in P), while the minimization problem is not; the decision version of the problem which determines whether a natural class can be defined with k features or less is NP-complete. We also show that, empirically, a greedy algorithm for finding minimal feature specifications will sometimes fail, and thus cannot be assumed to be the basis for human performance in solving the problem.

Demos

Time: 15:30–17:00

Location: Elite Hall B

CNNs for NLP in the Browser: Client-Side Deployment and Visualization Opportunities*Yiyun Liang, Zhucheng Tu, Laetitia Huang, and Jimmy Lin*

We demonstrate a JavaScript implementation of a convolutional neural network that performs feedforward inference completely in the browser. Such a deployment means that models can run completely on the client, on a wide range of devices, without making backend server requests. This design is useful for applications with stringent latency requirements or low connectivity. Our evaluations show the feasibility of JavaScript as a deployment target. Furthermore, an in-browser implementation enables seamless integration with the JavaScript ecosystem for information visualization, providing opportunities to visually inspect neural networks and better understand their inner workings.

Generating Continuous Representations of Medical Texts*Graham Spinks and Marie-Francine Moens*

We present an architecture that generates medical texts while learning an informative, continuous representation with discriminative features. During training the input to the system is a dataset of captions for medical X-Rays. The acquired continuous representations are of particular interest for use in many machine learning techniques where the discrete and high-dimensional nature of textual input is an obstacle. We use an Adversarially Regularized Autoencoder to create realistic text in both an unconditional and conditional setting. We show that this technique is applicable to medical texts which often contain syntactic and domain-specific shorthands. A quantitative evaluation shows that we achieve a lower model perplexity than a traditional LSTM generator.

Vis-Eval Metric Viewer: A Visualisation Tool for Inspecting and Evaluating Metric Scores of Machine Translation Output*David Steele and Lucia Specia*

Machine Translation systems are usually evaluated and compared using automated evaluation metrics such as BLEU and METEOR to score the generated translations against human translations. However, the interaction with the output from the metrics is relatively limited and results are commonly a single score along with a few additional statistics. Whilst this may be enough for system comparison it does not provide much useful feedback or a means for inspecting translations and their respective scores. VisEval Metric Viewer VEMV is a tool designed to provide visualisation of multiple evaluation scores so they can be easily interpreted by a user. VEMV takes in the source, reference, and hypothesis files as parameters, and scores the hypotheses using several popular evaluation metrics simultaneously. Scores are produced at both the sentence and dataset level and results are written locally to a series of HTML files that can be viewed on a web browser. The individual scored sentences can easily be inspected using powerful search and selection functions and results can be visualised with graphical representations of the scores and distributions.

Session 8 Overview – Sunday, June 3, 2018

Test-of-Time Paper Session (in memory of Aravind Joshi)

Empire B

BLEU: a Method for Automatic Evaluation of Machine Translation (ACL, 2002)

K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu

Discriminative Training Methods for Hidden Markov Models: Theory and Experiments with Perceptron Algorithms (EMNLP, 2002)

M. Collins

Thumbs up? Sentiment Classification using Machine Learning Techniques (EMNLP, 2002)

B. Pang, L. Lee, and S. Vaithyanathan

17:40

17:15

18:05

Session 8

Test-of-Time Paper Session (in memory of Aravind Joshi)

Empire B

Chair: Marilyn Walker

BLEU: a Method for Automatic Evaluation of Machine Translation (ACL, 2002)

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu

17:15–17:40

Human evaluations of machine translation are extensive but expensive. Human evaluations can take months to finish and involve human labor that can not be reused. We propose a method of automatic machine translation evaluation that is quick, inexpensive, and language-independent, that correlates highly with human evaluation, and that has little marginal cost per run. We present this method as an automated understudy to skilled human judges which substitutes for them when there is need for quick or frequent evaluations (so we call our method the bilingual evaluation understudy, BLEU).

Discriminative Training Methods for Hidden Markov Models: Theory and Experiments with Perceptron Algorithms (EMNLP, 2002)

Michael Collins

17:40–18:05

We describe new algorithms for training tagging models, as an alternative to maximum-entropy models or conditional random Fields (CRFs). The algorithms rely on Viterbi decoding of training examples, combined with simple additive updates. We describe theory justifying the algorithms through a modification of the proof of convergence of the perceptron algorithm for classification problems. We give experimental results on part-of-speech tagging and base noun phrase chunking, in both cases showing improvements over results for a maximum-entropy tagger.

Thumbs up? Sentiment Classification using Machine Learning Techniques (EMNLP, 2002)

Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan

18:05–18:30

We consider the problem of classifying documents not by topic, but by overall sentiment, e.g., determining whether a review is positive or negative. Using movie reviews as data, we find that standard machine learning techniques definitively outperform human-produced baselines. However, the three machine learning methods we employed (Naive Bayes, maximum entropy classification, and support vector machines) do not perform as well on sentiment classification as on traditional topic-based categorization. We conclude by examining factors that make the sentiment classification problem more challenging.

Social Event



Sunday, June 3, 2018, 18:45 – 21:00

Blaine Kern's Mardi Gras World
<https://www.mardigrasworld.com/>

Catch the shuttle bus outside of the Hyatt Regency New Orleans, conference hotel, at **18:45 sharp**. Bus loading will be to the back side of the escalator near the reservation desk. Please do not be late. This will take you the short ride to:

Blaine Kern's Mardi Gras World

Mardi Gras World is a world of wonders created for you by the people who bring Mardi Gras to life every year—the artists of Blaine Kern Studios. It's all in one magical place where you can peek behind the curtain and see Mardi Gras in the making.

Since 1947, Blaine Kern Studios has been as much a part of Carnival as the parades New Orleans loves. In fact, they create most of those parades, from concept through completion. They are **the world's leading makers of floats, sculpture and props**. Their work and props are located in theme parks, casinos and amusement parks around the globe—from here to Shanghai.

Your event **begins with a walk-through of the float den** among larger-than-life works of art. Visit The Studio, where you can watch artists designing, carving, painting and building for next year's parades. Guests will **explore props of every theme imaginable while enjoying passed hors d'oeuvres and a traditional New Orleans band, complete with a Grand Marshall**.

Guests will then follow stilt walkers to make their way to the **dockside River City Plaza** and **River City Ballroom** for New Orleans' famous cuisine and libations and live **Zydeco, funk, soul and R&B from one of the best performers in the city**. This will surely be a night to remember!

Main Conference: Monday, June 4

Overview

07:30 – 18:00	Registration	<i>Empire Foyer</i>			
07:45 – 08:45	Breakfast	<i>Empire Foyer</i>			
08:45 – 09:00	Announcements	<i>Empire Ballroom</i>			
09:00 – 10:00	Keynote Address: Dilek Hakkani-Tür (sponsored by Bloomberg)	<i>Empire Ballroom</i>			
10:00 – 10:30	Morning Coffee	<i>Empire Foyer</i>			
	Research Oral 9	Poster/Demo 9/10 (10:30-12:00)			
10:30 – 11:30	Information Extraction 4 <i>Empire A</i>	Semantics 4 <i>Empire B</i>	Generation 2 <i>Empire C</i>	Question Answering 2 / Social Media & Computational Social Science 3 / Summarization 2 / Demos	
	Research Oral 10	Industry Oral 10			
11:30 – 12:30	Dialogue and Interactive Systems 1 <i>Empire A</i>	Information Extraction 5 <i>Empire B</i>	Generation 3 <i>Empire C</i>	Machine Learning - Prediction <i>Empire D</i>	<i>Elite Hall B</i>
12:30 – 14:00	Lunch				
13:00 – 14:00	NAACL Business Meeting				
	Research Oral 11			Poster/Demo 11	
14:00 – 15:30	Sentiment Analysis 2 <i>Empire A</i>	Discourse and Pragmatics 2 <i>Empire B</i>	Tagging, Chunking, Syntax and Parsing 3 <i>Empire C</i>	Cognitive Modeling and Psycholinguistics 2 / Dialogue and Interactive Systems 2 / Text Mining 2 / Speech 2 / Vision, Robotics and Other Grounding 3 / Demos <i>Elite Hall B</i>	
15:30 – 16:00	Afternoon Coffee				<i>Empire Foyer</i>
17:00 – 18:15	Outstanding Papers and Closing Session				<i>Empire Ballroom</i>

Keynote Address: Dilek Hakkani-Tür (sponsored by Bloomberg)

Google Assistant or My Assistant? Towards Personalized Situated Conversational Agents

Monday, June 4, 2018, 9:00–10:10

Chair: Heng Ji

Empire Ballroom

Abstract: Interacting with machines in natural language has been a holy grail since the beginning of computers. Given the difficulty of understanding natural language, only in the past couple of decades, we started seeing real user applications for targeted/limited domains. More recently, advances in deep learning based approaches enabled exciting new research frontiers for end-to-end goal-oriented conversational systems. However, personalization (i.e., learning to take actions from users and learning about users beyond memorizing simple attributes) remains a research challenge. In this talk, I'll review end-to-end situated dialogue systems research, with components for situated language understanding, dialogue state tracking, policy, and language generation. The talk will highlight novel approaches where dialogue is viewed as a collaborative game between a user and an agent in the presence of visual information. The situated conversational agent can be bootstrapped using user simulation (crawl), improved through interactions with crowd-workers (walk), and iteratively refined with real user interactions (run).

Biography: Dilek is a research scientist at Google Research Dialogue Group and has previously held positions at Microsoft Research, ICSI, and AT&T Labs – Research. She is a fellow of the IEEE and of ISCA. Her research interests include conversational AI, natural language and speech processing, spoken dialogue systems, and machine learning for language processing.

Session 9 Overview – Monday, June 4, 2018

Track A <i>Information Extraction 4</i> Empire A	Track B <i>Semantics 4</i> Empire B	Track C <i>Generation 2</i> Empire C	Track P <i>Posters and Demos (10:30–12:00)</i> Elite Hall B
A Neural Layered Model for Nested Named Entity Recognition <i>M. Ju, M. Miwa, and S. Ananiadou</i>	Multimodal Frame Identification with Multilingual Evaluation <i>T. Botschen, I. Gurevych, J.-C. Klie, H. Mousselly Sergieh, and S. Roth</i>	Bootstrapping Generators from Noisy Data <i>L. Perez-Beltrachini and M. Lapata</i>	Question Answering 2 / Social Media & Computational Social Science 3 / Summarization 2 / Demos
DR-BiLSTM: Dependent Reading Bidirectional LSTM for Natural Language Inference <i>R. Ghaeini, S. A. Hasan, V. Datla, J. Liu, K. Lee, A. Qadir, Y. Ling, A. Prakash, X. Fern, and O. Farri</i>	Learning Joint Semantic Parsers from Disjoint Data <i>H. Peng, S. Thomson, S. Swayamdipta, and N. A. Smith</i>	SHAPED: Shared-Private Encoder-Decoder for Text Style Adaptation <i>Y. Zhang, N. Ding, and R. Soricut</i>	10:30
KBGAN: Adversarial Learning for Knowledge Graph Embeddings <i>L. Cai and W. Y. Wang</i>	Identifying Semantic Divergences in Parallel Text without Annotations <i>Y. Vyas, X. Niu, and M. Carpuat</i>	Generating Descriptions from Structured Data Using a Bifocal Attention Mechanism and Gated Orthogonalization <i>P. Nema, S. Shetty, P. Jain, A. Laha, K. Sankaranarayanan, and M. M. Khapra</i>	10:48 11:06

Parallel Session 9

Information Extraction 4

Empire A

Chair: Bonan Min

A Neural Layered Model for Nested Named Entity Recognition

Meizhi Ju, Makoto Miwa, and Sophia Ananiadou

10:30–10:47

Entity mentions embedded in longer entity mentions are referred to as nested entities. Most named entity recognition (NER) systems deal only with the flat entities and ignore the inner nested ones, which fails to capture finer-grained semantic information in underlying texts. To address this issue, we propose a novel neural model to identify nested entities by dynamically stacking flat NER layers. Each flat NER layer is based on the state-of-the-art flat NER model that captures sequential context representation with bidirectional Long Short-Term Memory (LSTM) layer and feeds it to the cascaded CRF layer. Our model merges the output of the LSTM layer in the current flat NER layer to build new representation for detected entities and subsequently feeds them into the next flat NER layer. This allows our model to extract outer entities by taking full advantage of information encoded in their corresponding inner entities, in an inside-to-outside way. Our model dynamically stacks the flat NER layers until no outer entities are extracted. Extensive evaluation shows that our dynamic model outperforms state-of-the-art feature-based systems on nested NER, achieving 74.7% and 72.2% on GENIA and ACE2005 datasets, respectively, in terms of F-score.

DR-BiLSTM: Dependent Reading Bidirectional LSTM for Natural Language Inference

Reza Ghaeini, Sadid A. Hasan, Vivek Datta, Joey Liu, Kathy Lee, Ashequl Qadir, Yuan Ling, Aditya Prakash, Xiaoli Fern, and Oladimeji Farri

10:48–11:05

We present a novel deep learning architecture to address the natural language inference (NLI) task. Existing approaches mostly rely on simple reading mechanisms for independent encoding of the premise and hypothesis. Instead, we propose a novel dependent reading bidirectional LSTM network (DR-BiLSTM) to efficiently model the relationship between a premise and a hypothesis during encoding and inference. We also introduce a sophisticated ensemble strategy to combine our proposed models, which noticeably improves final predictions. Finally, we demonstrate how the results can be improved further with an additional preprocessing step. Our evaluation shows that DR-BiLSTM obtains the best single model and ensemble model results achieving the new state-of-the-art scores on the Stanford NLI dataset.

KBGAN: Adversarial Learning for Knowledge Graph Embeddings

Liwei Cai and William Yang Wang

11:06–11:23

We introduce KBGAN, an adversarial learning framework to improve the performances of a wide range of existing knowledge graph embedding models. Because knowledge graphs typically only contain positive facts, sampling useful negative training examples is a nontrivial task. Replacing the head or tail entity of a fact with a uniformly randomly selected entity is a conventional method for generating negative facts, but the majority of the generated negative facts can be easily discriminated from positive facts, and will contribute little towards the training. Inspired by generative adversarial networks (GANs), we use one knowledge graph embedding model as a negative sample generator to assist the training of our desired model, which acts as the discriminator in GANs. This framework is independent of the concrete form of generator and discriminator, and therefore can utilize a wide variety of knowledge graph embedding models as its building blocks. In experiments, we adversarially train two translation-based models, TRANSE and TRANSD, each with assistance from one of the two probability-based models, DISTMULT and COMPLEX. We evaluate the performances of KBGAN on the link prediction task, using three knowledge base completion datasets: FB15k-237, WN18 and WN18RR. Experimental results show that adversarial training substantially improves the performances of target embedding models under various settings.

Semantics 4

Empire B

Chair: Yoav Artzi

Multimodal Frame Identification with Multilingual Evaluation

Teresa Botschen, Iryna Gurevych, Jan-Christoph Klie, Hatem Mousselly Sergieh, and Stefan Roth

10:30–10:47

An essential step in FrameNet Semantic Role Labeling is the Frame Identification (FrameId) task, which aims at disambiguating a situation around a predicate. Whilst current FrameId methods rely on textual representations only, we hypothesize that FrameId can profit from a richer understanding of the situational context. Such contextual information can be obtained from common sense knowledge, which is more present in images than in text. In this paper, we extend a state-of-the-art FrameId system in order to effectively leverage multimodal representations. We conduct a comprehensive evaluation on the English FrameNet and its German counterpart SALSA. Our analysis shows that for the German data, textual representations are still competitive with multimodal ones. However on the English data, our multimodal FrameId approach outperforms its unimodal counterpart, setting a new state of the art. Its benefits are particularly apparent in dealing with ambiguous and rare instances, the main source of errors of current systems. For research purposes, we release (a) the implementation of our system, (b) our evaluation splits for SALSA 2.0, and (c) the embeddings for synsets and IMAGINED words.

Learning Joint Semantic Parsers from Disjoint Data

Hao Peng, Sam Thomson, Swabha Swayamdipta, and Noah A. Smith

10:48–11:05

We present a new approach to learning a semantic parser from multiple datasets, even when the target semantic formalisms are drastically different and the underlying corpora do not overlap. We handle such “disjoint” data by treating annotations for unobserved formalisms as latent structured variables. Building on state-of-the-art baselines, we show improvements both in frame-semantic parsing and semantic dependency parsing by modeling them jointly.

Identifying Semantic Divergences in Parallel Text without Annotations

Yogarshi Vyas, Xing Niu, and Marine Carpuat

11:06–11:23

Recognizing that even correct translations are not always semantically equivalent, we automatically detect meaning divergences in parallel sentence pairs with a deep neural model of bilingual semantic similarity which can be trained for any parallel corpus without any manual annotation. We show that our semantic model detects divergences more accurately than models based on surface features derived from word alignments, and that these divergences matter for neural machine translation.

Generation 2

Empire C

Chair: Michael White

Bootstrapping Generators from Noisy Data

Laura Perez-Beltrachini and Mirella Lapata

10:30–10:47

A core step in statistical data-to-text generation concerns learning correspondences between structured data representations (e.g., facts in a database) and associated texts. In this paper we aim to bootstrap generators from large scale datasets where the data (e.g., DBpedia facts) and related texts (e.g., Wikipedia abstracts) are loosely aligned. We tackle this challenging task by introducing a special-purpose content selection mechanism. We use multi-instance learning to automatically discover correspondences between data and text pairs and show how these can be used to enhance the content signal while training an encoder-decoder architecture. Experimental results demonstrate that models trained with content-specific objectives improve upon a vanilla encoder-decoder which solely relies on soft attention.

SHAPED: Shared-Private Encoder-Decoder for Text Style Adaptation

Ye Zhang, Nan Ding, and Radu Soricut

10:48–11:05

Supervised training of abstractive language generation models results in learning conditional probabilities over language sequences based on the supervised training signal. When the training signal contains a variety of writing styles, such models may end up learning an ‘average’ style that is directly influenced by the training data make-up and cannot be controlled by the needs of an application. We describe a family of model architectures capable of capturing both generic language characteristics via shared model parameters, as well as particular style characteristics via private model parameters. Such models are able to generate language according to a specific learned style, while still taking advantage of their power to model generic language phenomena. Furthermore, we describe an extension that uses a mixture of output distributions from all learned styles to perform on-the-fly style adaptation based on the textual input alone. Experimentally, we find that the proposed models consistently outperform models that encapsulate single-style or average-style language generation capabilities.

Generating Descriptions from Structured Data Using a Bifocal Attention Mechanism and Gated Orthogonalization

Preksha Nema, Shreyas Shetty, Parag Jain, Anirban Laha, Karthik Sankaranarayanan, and Mitesh M. Khapra

11:06–11:23

In this work, we focus on the task of generating natural language descriptions from a structured table of facts containing fields (such as nationality, occupation, etc) and values (such as Indian, actor, director, etc). One simple choice is to treat the table as a sequence of fields and values and then use a standard seq2seq model for this task. However, such a model is too generic and does not exploit task specific characteristics. For example, while generating descriptions from a table, a human would attend to information at two levels: (i) the fields (macro level) and (ii) the values within the field (micro level). Further, a human would continue attending to a field for a few timesteps till all the information from that field has been rendered and then never return back to this field (because there is nothing left to say about it). To capture this behavior we use (i) a fused bifocal attention mechanism which exploits and combines this micro and macro level information and (ii) a gated orthogonalization mechanism which tries to ensure that a field is remembered for a few time steps and then forgotten. We experiment with a recently released dataset which contains fact tables about people and their corresponding one line biographical descriptions in English. In addition, we also introduce two similar datasets for French and German. Our experiments show that the proposed model gives 21% relative improvement over a recently proposed state of the art method and 10% relative improvement over basic seq2seq models. The code and the datasets developed as a part of this work are publicly available on https://github.com/PrekshaNema25/StructuredData_To_Descriptions

Session 10 Overview – Monday, June 4, 2018

Track A <i>Dialogue and Interactive Systems I</i>	Track B <i>Information Extraction 5</i>	Track C <i>Generation 3</i>	Track D <i>Industry: Machine Learning - Prediction</i>	Track P <i>Posters and Demos (10:30–12:00)</i>
Empire A A Hierarchical Latent Structure for Variational Conversation Modeling <i>Y. Park, J. Cho, and G. Kim</i>	Empire B Variational Knowledge Graph Reasoning <i>W. Chen, W. Xiong, X. Yan, and W. Y. Wang</i>	Empire C Interpretable Charge Predictions for Criminal Cases: Learning to Generate Court Views from Fact Descriptions <i>H. Ye, X. Jiang, Z. Luo, and W. Chao</i>	Empire D Prediction for the Newsroom: Which Articles Will Get the Most Comments? (INDUSTRY) <i>C. Ambroselli, J. Risch, R. Krestel, and A. Loos</i>	Elite Hall B Question Answering 2 / Social Media & Computational Social Science 3 / Summarization 2 / Demos 11:30
Detecting Egregious Conversations between Customers and Virtual Agents <i>T. Sandbank, M. Shmueli-Scheuer, J. Herzog, D. Konopnicki, J. Richards, and D. Piorkowski</i>	Inducing Temporal Relations from Time Anchor Annotation <i>F. Cheng and Y. Miyao</i>	Delete, Retrieve, Generate: a Simple Approach to Sentiment and Style Transfer <i>J. Li, R. Jia, H. He, and P. Liang</i>	Demand-Weighted Completeness Prediction for a Knowledge Base (INDUSTRY) <i>A. Hopkinson, A. Gurdasani, D. Palfray, and A. Mittal</i>	Question Answering 2 / Social Media & Computational Social Science 3 / Summarization 2 / Demos 11:48
Learning to Disentangle Interleaved Conversational Threads with a Siamese Hierarchical Network and Similarity Ranking <i>J.-Y. Jiang, F. Chen, Y.-Y. Chen, and W. Wang</i>	ELDEN: Improved Entity Linking Using Densified Knowledge Graphs <i>P. Radhakrishnan, P. Talukdar, and V. Varma</i>	Adversarial Example Generation with Syntactically Controlled Paraphrase Networks <i>M. Iyyer, J. Wieting, K. Gimpel, and L. Zettlemoyer</i>	Personalized Neural Language Models for Real-World Query Autocompletion (INDUSTRY) <i>N. Fiorini and Z. Lu</i>	12:06

Parallel Session 10

Oral: Dialogue and Interactive Systems 1

Empire A

Chair: Yun-Nung Chen

A Hierarchical Latent Structure for Variational Conversation Modeling

Yookoon Park, Jaemin Cho, and Gunhee Kim

11:30–11:47

Variational autoencoders (VAE) combined with hierarchical RNNs have emerged as a powerful framework for conversation modeling. However, they suffer from the notorious degeneration problem, where the decoders learn to ignore latent variables and reduce to vanilla RNNs. We empirically show that this degeneracy occurs mostly due to two reasons. First, the expressive power of hierarchical RNN decoders is often high enough to model the data using only its decoding distributions without relying on the latent variables. Second, the conditional VAE structure whose generation process is conditioned on a context, makes the range of training targets very sparse; that is, the RNN decoders can easily overfit to the training data ignoring the latent variables. To solve the degeneration problem, we propose a novel model named Variational Hierarchical Conversation RNNs (VHCR), involving two key ideas of (1) using a hierarchical structure of latent variables, and (2) exploiting an utterance drop regularization. With evaluations on two datasets of Cornell Movie Dialog and Ubuntu Dialog Corpus, we show that our VHCR successfully utilizes latent variables and outperforms state-of-the-art models for conversation generation. Moreover, it can perform several new utterance control tasks, thanks to its hierarchical latent structure.

Detecting Egregious Conversations between Customers and Virtual Agents

Tommy Sandbank, Michal Shmueli-Scheuer, Jonathan Herzig, David Konopnicki, John Richards, and David Piorkowski

11:48–12:05

Virtual agents are becoming a prominent channel of interaction in customer service. Not all customer interactions are smooth, however, and some can become almost comically bad. In such instances, a human agent might need to step in and salvage the conversation. Detecting bad conversations is important since disappointing customer service may threaten customer loyalty and impact revenue. In this paper, we outline an approach to detecting such egregious conversations, using behavioral cues from the user, patterns in agent responses, and user-agent interaction. Using logs of two commercial systems, we show that using these features improves the detection F1-score by around 20% over using textual features alone. In addition, we show that those features are common across two quite different domains and, arguably, universal.

Learning to Disentangle Interleaved Conversational Threads with a Siamese Hierarchical Network and Similarity Ranking

Jyun-Yu Jiang, Francine Chen, Yan-Ying Chen, and Wei Wang

12:06–12:23

An enormous amount of conversation occurs online every day, such as on chat platforms where multiple conversations may take place concurrently. Interleaved conversations lead to difficulties in not only following discussions but also retrieving relevant information from simultaneous messages. Conversation disentanglement aims to separate intermingled messages into detached conversations. In this paper, we propose to leverage representation learning for conversation disentanglement. A Siamese hierarchical convolutional neural network (SHCNN), which integrates local and more global representations of a message, is first presented to estimate the conversation-level similarity between closely posted messages. With the estimated similarity scores, our algorithm for conversation identification by similarity ranking (CISIR) then derives conversations based on high-confidence message pairs and pairwise redundancy. Experiments were conducted with four publicly available datasets of conversations from Reddit and IRC channels. The experimental results show that our approach significantly outperforms comparative baselines in both pairwise similarity estimation and conversation disentanglement.

Oral: Information Extraction 5

Empire B

Chair: Xiang Ren

Variational Knowledge Graph Reasoning

Wenhu Chen, Wenhan Xiong, Xifeng Yan, and William Yang Wang

11:30–11:47

Inferring missing links in knowledge graphs (KG) has attracted a lot of attention from the research community. In this paper, we tackle a practical query answering task involving predicting the relation of a given entity

pair. We frame this prediction problem as an inference problem in a probabilistic graphical model and aim at resolving it from a variational inference perspective. In order to model the relation between the query entity pair, we assume that there exists an underlying latent variable (paths connecting two nodes) in the KG, which carries the equivalent semantics of their relations. However, due to the intractability of connections in large KGs, we propose to use variation inference to maximize the evidence lower bound. More specifically, our framework (Diva) is composed of three modules, i.e. a posterior approximator, a prior (path finder), and a likelihood (path reasoner). By using variational inference, we are able to incorporate them closely into a unified architecture and jointly optimize them to perform KG reasoning. With active interactions among these sub-modules, DIVA is better at handling noise and coping with more complex reasoning scenarios. In order to evaluate our method, we conduct the experiment of the link prediction task on multiple datasets and achieve state-of-the-art performances on both datasets.

Inducing Temporal Relations from Time Anchor Annotation

Fei Cheng and Yusuke Miyao

11:48–12:05

Recognizing temporal relations among events and time expressions has been an essential but challenging task in natural language processing. Conventional annotation of judging temporal relations puts a heavy load on annotators. In reality, the existing annotated corpora include annotations on only "salient" event pairs, or on pairs in a fixed window of sentences. In this paper, we propose a new approach to obtain temporal relations from absolute time value (a.k.a. time anchors), which is suitable for texts containing rich temporal information such as news articles. We start from time anchors for events and time expressions, and temporal relation annotations are induced automatically by computing relative order of two time anchors. This proposal shows several advantages over the current methods for temporal relation annotation: it requires less annotation effort, can induce inter-sentence relations easily, and increases informativeness of temporal relations. We compare the empirical statistics and automatic recognition results with our data against a previous temporal relation corpus. We also reveal that our data contributes to a significant improvement of the downstream time anchor prediction task, demonstrating 14.1 point increase in overall accuracy.

ELDEN: Improved Entity Linking Using Densified Knowledge Graphs

Priya Radhakrishnan, Partha Talukdar, and Vasudeva Varma

12:06–12:23

Entity Linking (EL) systems aim to automatically map mentions of an entity in text to the corresponding entity in a Knowledge Graph (KG). Degree of connectivity of an entity in the KG directly affects an EL system's ability to correctly link mentions in text to the entity in KG. This causes many EL systems to perform well for entities well connected to other entities in KG, bringing into focus the role of KG density in EL. In this paper, we propose Entity Linking using Densified Knowledge Graphs (ELDEN). ELDEN is an EL system which first densifies the KG with co-occurrence statistics from a large text corpus, and then uses the densified KG to train entity embeddings. Entity similarity measured using these trained entity embeddings result in improved EL. ELDEN outperforms state-of-the-art EL system on benchmark datasets. Due to such densification, ELDEN performs well for sparsely connected entities in the KG too. ELDEN's approach is simple, yet effective. We have made ELDEN's code and data publicly available.

Oral: Generation 3

Empire C

Chair: Amanda Stent

Interpretable Charge Predictions for Criminal Cases: Learning to Generate Court Views from Fact Descriptions

Hai Ye, Xin Jiang, Zhunchen Luo, and Wenhan Chao

11:30–11:47

In this paper, we propose to study the problem of court view generation from the fact description in a criminal case. The task aims to improve the interpretability of charge prediction systems and help automatic legal document generation. We formulate this task as a text-to-text natural language generation (NLG) problem. Sequence-to-sequence model has achieved cutting-edge performances in many NLG tasks. However, due to the non-distinctions of fact descriptions, it is hard for Seq2Seq model to generate charge-discriminative court views. In this work, we explore charge labels to tackle this issue. We propose a label-conditioned Seq2Seq model with attention for this problem, to decode court views conditioned on encoded charge labels. Experimental results show the effectiveness of our method.

Delete, Retrieve, Generate: a Simple Approach to Sentiment and Style Transfer

Juncen Li, Robin Jia, He He, and Percy Liang

11:48–12:05

We consider the task of text attribute transfer: transforming a sentence to alter a specific attribute (e.g., sentiment) while preserving its attribute-independent content (e.g., “screen is just the right size” to “screen is too small”). Our training data includes only sentences labeled with their attribute (e.g., positive and negative), but not pairs of sentences that only differ in the attributes, so we must learn to disentangle attributes from attribute-independent content in an unsupervised way. Previous work using adversarial methods has struggled to produce high-quality outputs. In this paper, we propose simpler methods motivated by the observation that text attributes are often marked by distinctive phrases (e.g., “too small”). Our strongest method extracts content words by deleting phrases associated with the sentence’s original attribute value, retrieves new phrases associated with the target attribute, and uses a neural model to fluently combine these into a final output. Based on human evaluation, our best method generates grammatical and appropriate responses on 22% more inputs than the best previous system, averaged over three attribute transfer datasets: altering sentiment of reviews on Yelp, altering sentiment of reviews on Amazon, and altering image captions to be more romantic or humorous.

Adversarial Example Generation with Syntactically Controlled Paraphrase Networks

Mohit Iyyer, John Wieting, Kevin Gimpel, and Luke Zettlemoyer

12:06–12:23

We propose syntactically controlled paraphrase networks (SCPNs) and use them to generate adversarial examples. Given a sentence and a target syntactic form (e.g., a constituency parse), SCPNs are trained to produce a paraphrase of the sentence with the desired syntax. We show it is possible to create training data for this task by first doing backtranslation at a very large scale, and then using a parser to label the syntactic transformations that naturally occur during this process. Such data allows us to train a neural encoder-decoder model with extra inputs to specify the target syntax. A combination of automated and human evaluations show that SCPNs generate paraphrases that follow their target specifications without decreasing paraphrase quality when compared to baseline (uncontrolled) paraphrase systems. Furthermore, they are more capable of generating syntactically adversarial examples that both (1) “fool” pretrained models and (2) improve the robustness of these models to syntactic variation when used to augment their training data.

Industry: Machine Learning - Prediction

Empire D

Chair: Gokhan Tür

Prediction for the Newsroom: Which Articles Will Get the Most Comments? (INDUSTRY)

Carl Ambroselli, Julian Risch, Ralf Krestel, and Andreas Loos

11:30–11:47

The overwhelming success of the Web and mobile technologies has enabled millions to share their opinions publicly at any time. But the same success also endangers this freedom of speech due to closing down of participatory sites misused by individuals or interest groups. We propose to support manual moderation by proactively drawing the attention of our moderators to article discussions that most likely need their intervention. To this end, we predict which articles will receive a high number of comments. In contrast to existing work, we enrich the article with metadata, extract semantic and linguistic features, and exploit annotated data from a foreign language corpus. Our logistic regression model improves F1-scores by over 80% in comparison to state-of-the-art approaches.

Demand-Weighted Completeness Prediction for a Knowledge Base (INDUSTRY)

Andrew Hopkinson, Amit Gurdasani, Dave Palfrey, and Arpit Mittal

11:48–12:05

In this paper we introduce the notion of Demand-Weighted Completeness, allowing estimation of the completeness of a knowledge base with respect to how it is used. Defining an entity by its classes, we employ usage data to predict the distribution over relations for that entity. For example, instances of person in a knowledge base may require a birth date, name and nationality to be considered complete. These predicted relation distributions enable detection of important gaps in the knowledge base, and define the required facts for unseen entities. Such characterisation of the knowledge base can also quantify how usage and completeness change over time. We demonstrate a method to measure Demand-Weighted Completeness, and show that a simple neural network model performs well at this prediction task.

Personalized Neural Language Models for Real-World Query Autocompletion (INDUSTRY)

Nicolas Fiorini and Zhiyong Lu

12:06–12:23

Query auto completion (QAC) systems are a standard part of search engines in industry, helping users formulate their query. Such systems update their suggestions after the user types each character, predicting the user’s

intent using various signals — one of the most common being popularity. Recently, deep learning approaches have been proposed for the QAC task, to specifically address the main limitation of previous popularity-based methods: the inability to predict unseen queries. In this work we improve previous methods based on neural language modeling, with the goal of building an end-to-end system. We particularly focus on using real-world data by integrating user information for personalized suggestions when possible. We also make use of time information and study how to increase diversity in the suggestions while studying the impact on scalability. Our empirical results demonstrate a marked improvement on two separate datasets over previous best methods in both accuracy and scalability, making a step towards neural query auto-completion in production search engines.

Posters and Demos: Sessions 9/10

Posters: Question Answering 2

Time: 10:30–12:00

Location: Elite Hall B

Chair: Lu Wang

CliCR: a Dataset of Clinical Case Reports for Machine Reading Comprehension

Simon Suster and Walter Daelemans

We present a new dataset for machine comprehension in the medical domain. Our dataset uses clinical case reports with around 100,000 gap-filling queries about these cases. We apply several baselines and state-of-the-art neural readers to the dataset, and observe a considerable gap in performance (20% F1) between the best human and machine readers. We analyze the skills required for successful answering and show how reader performance varies depending on the applicable skills. We find that inferences using domain knowledge and object tracking are the most frequently required skills, and that recognizing omitted information and spatio-temporal reasoning are the most difficult for the machines.

Contextualized Word Representations for Reading Comprehension

Shimi Salant and Jonathan Berant

Reading a document and extracting an answer to a question about its content has attracted substantial attention recently. While most work has focused on the interaction between the question and the document, in this work we evaluate the importance of context when the question and document are processed independently. We take a standard neural architecture for this task, and show that by providing rich contextualized word representations from a large pre-trained language model as well as allowing the model to choose between context-dependent and context-independent word representations, we can obtain dramatic improvements and reach performance comparable to state-of-the-art on the competitive SQuAD dataset.

Crowdsourcing Question-Answer Meaning Representations

Julian Michael, Gabriel Stanovsky, Luheng He, Ido Dagan, and Luke Zettlemoyer

We introduce Question-Answer Meaning Representations (QAMRs), which represent the predicate-argument structure of a sentence as a set of question-answer pairs. We develop a crowdsourcing scheme to show that QAMRs can be labeled with very little training, and gather a dataset with over 5,000 sentences and 100,000 questions. A qualitative analysis demonstrates that the crowd-generated question-answer pairs cover the vast majority of predicate-argument relationships in existing datasets (including PropBank, NomBank, and QA-SRL) along with many previously under-resourced ones, including implicit arguments and relations. We also report baseline models for question generation and answering, and summarize a recent approach for using QAMR labels to improve an Open IE system. These results suggest the freely available QAMR data and annotation scheme should support significant future work.

Learning to Collaborate for Question Answering and Asking

Duyu Tang, Nan Duan, Zhao Yan, Zhirui Zhang, Yibo Sun, Shujie Liu, Yuanhua Lv, and Ming Zhou

Question answering (QA) and question generation (QG) are closely related tasks that could improve each other; however, the connection of these two tasks is not well explored in literature. In this paper, we give a systematic study that seeks to leverage the connection to improve both QA and QG. We present a training algorithm that generalizes both Generative Adversarial Network (GAN) and Generative Domain-Adaptive Nets (GDAN) under the question answering scenario. The two key ideas are improving the QG model with QA through incorporating additional QA-specific signal as the loss function, and improving the QA model with QG through adding artificially generated training instances. We conduct experiments on both document based and knowledge based question answering tasks. We have two main findings. Firstly, the performance of a QG model (e.g. in terms of BLEU score) could be easily improved by a QA model via policy gradient. Secondly, directly applying GAN that regards all the generated questions as negative instances could not improve the accuracy of the QA model. Learning when to regard generated questions as positive instances could bring performance boost.

Learning to Rank Question-Answer Pairs Using Hierarchical Recurrent Encoder with Latent Topic Clustering*Seunghyun Yoon, Joongbo Shin, and Kyomin Jung*

In this paper, we propose a novel end-to-end neural architecture for ranking candidate answers, that adapts a hierarchical recurrent neural network and a latent topic clustering module. With our proposed model, a text is encoded to a vector representation from a word-level to a chunk-level to effectively capture the entire meaning. In particular, by adapting the hierarchical structure, our model shows very small performance degradations in longer text comprehension while other state-of-the-art recurrent neural network models suffer from it. Additionally, the latent topic clustering module extracts semantic information from target samples. This clustering module is useful for any text related tasks by allowing each data sample to find its nearest topic cluster, thus helping the neural network model analyze the entire data. We evaluate our models on the Ubuntu Dialogue Corpus and consumer electronic domain question answering dataset, which is related to Samsung products. The proposed model shows state-of-the-art results for ranking question-answer pairs.

Leveraging Context Information for Natural Question Generation*Linfeng Song, Zhiguo Wang, Wael Hamza, Yue Zhang, and Daniel Gildea*

The task of natural question generation is to generate a corresponding question given the input passage (fact) and answer. It is useful for enlarging the training set of QA systems. Previous work has adopted sequence-to-sequence models that take a passage with an additional bit to indicate answer position as input. However, they do not explicitly model the information between answer and other context within the passage. We propose a model that matches the answer with the passage before generating the question. Experiments show that our model outperforms the existing state of the art using rich features.

Robust Machine Comprehension Models via Adversarial Training*Yicheng Wang and Mohit Bansal*

It is shown that many published models for the Stanford Question Answering Dataset (Rajpurkar et al., 2016) lack robustness, suffering an over 50% decrease in F1 score during adversarial evaluation based on the AddSent (Jia and Liang, 2017) algorithm. It has also been shown that retraining models on data generated by AddSent has limited effect on their robustness. We propose a novel alternative adversary-generation algorithm, AddSentDiverse, that significantly increases the variance within the adversarial training data by providing effective examples that punish the model for making certain superficial assumptions. Further, in order to improve robustness to AddSent’s semantic perturbations (e.g., antonyms), we jointly improve the model’s semantic-relationship learning capabilities in addition to our AddSentDiverse-based adversarial training data augmentation. With these additions, we show that we can make a state-of-the-art model significantly more robust, achieving a 36.5% increase in F1 score under many different types of adversarial evaluation while maintaining performance on the regular SQuAD task.

Simple and Effective Semi-Supervised Question Answering*Bhuwan Dhingra, Danish Pruthi, and Dheeraj Rajagopal*

Recent success of deep learning models for the task of extractive Question Answering (QA) is hinged on the availability of large annotated corpora. However, large domain specific annotated corpora are limited and expensive to construct. In this work, we envision a system where the end user specifies a set of base documents and only a few labelled examples. Our system exploits the document structure to create cloze-style questions from these base documents; pre-trains a powerful neural network on the cloze style questions; and further fine-tunes the model on the labeled examples. We evaluate our proposed system across three diverse datasets from different domains, and find it to be highly effective with very little labeled data. We attain more than 50% F1 score on SQuAD and TriviaQA with less than a thousand labelled examples. We are also releasing a set of 3.2M cloze-style questions for practitioners to use while building QA systems.

Supervised and Unsupervised Transfer Learning for Question Answering*Yu-An Chung, Hung-yi Lee, and James Glass*

Although transfer learning has been shown to be successful for tasks like object and speech recognition, its applicability to question answering (QA) has yet to be well-studied. In this paper, we conduct extensive experiments to investigate the transferability of knowledge learned from a source QA dataset to a target dataset using two QA models. The performance of both models on a TOEFL listening comprehension test (Tseng et al., 2016) and MCTest (Richardson et al., 2013) is significantly improved via a simple transfer learning technique from MovieQA (Tapaswi et al., 2016). In particular, one of the models achieves the state-of-the-art on all target datasets; for the TOEFL listening comprehension test, it outperforms the previous best model by 7%. Finally, we show that transfer learning is helpful even in unsupervised scenarios when correct answers for

target QA dataset examples are not available.

Tracking State Changes in Procedural Text: a Challenge Dataset and Models for Process Paragraph Comprehension

Bhavana Dalvi, Lifu Huang, Niket Tandon, Wen-tau Yih, and Peter Clark

We present a new dataset and models for comprehending paragraphs about processes (e.g., photosynthesis), an important genre of text describing a dynamic world. The new dataset, ProPara, is the first to contain natural (rather than machine-generated) text about a changing world along with a full annotation of entity states (location and existence) during those changes (81k datapoints). The end-task, tracking the location and existence of entities through the text, is challenging because the causal effects of actions are often implicit and need to be inferred. We find that previous models that have worked well on synthetic data achieve only mediocre performance on ProPara, and introduce two new neural models that exploit alternative mechanisms for state prediction, in particular using LSTM input encoding and span prediction. The new models improve accuracy by up to 19%. We are releasing the ProPara dataset and our models to the community.

TypeSQL: Knowledge-Based Type-Aware Neural Text-to-SQL Generation

Tao Yu, Zifan Li, Zilin Zhang, Rui Zhang, and Dragomir Radev

Interacting with relational databases through natural language helps users with any background easily query and analyze a vast amount of data. This requires a system that understands users' questions and converts them to SQL queries automatically. In this paper, we present a novel approach TypeSQL which formats the problem as a slot filling task in a more reasonable way. In addition, TypeSQL utilizes type information to better understand rare entities and numbers in the questions. We experiment this idea on the WikiSQL dataset and outperform the prior art by 6% in much shorter time. We also show that accessing the content of databases can significantly improve the performance when users' queries are not well-formed. TypeSQL can reach 82.6% accuracy, a 17.5% absolute improvement compared to the previous content-sensitive model.

Training a Ranking Function for Open-Domain Question Answering (SRW)

Phu Mon Htut, Samuel Bowman, and Kyunghyun Cho

In recent years, there have been amazing advances in deep learning methods for machine reading. In machine reading, the machine reader has to extract the answer from the given ground truth paragraph. Recently, the state-of-the-art machine reading models achieve human level performance in SQuAD which is a reading comprehension-style question answering (QA) task. The success of machine reading has inspired researchers to combine Information Retrieval with machine reading to tackle open-domain QA. However, these systems perform poorly compared to reading comprehension-style QA because it is difficult to retrieve the pieces of paragraphs that contain the answer to the question. In this study, we propose two neural network rankers that assign scores to different passages based on their likelihood of containing the answer to a given question. Additionally, we analyze the relative importance of semantic similarity and word level relevance matching in open-domain QA.

Posters: Social Media and Computational Social Science 3

Time: 10:30–12:00

Location: Elite Hall B

Chair: Lu Wang

Combining Deep Learning and Topic Modeling for Review Understanding in Context-Aware Recommendation

Mingmin Jin, Xin Luo, Huiling Zhu, and Hankui Zhuo

With the rise of e-commerce, people are accustomed to writing their reviews after receiving the goods. These comments are so important that a bad review can have a direct impact on others buying. Besides, the abundant information within user reviews is very useful for extracting user preferences and item properties. In this paper, we investigate the approach to effectively utilize review information for recommender systems. The proposed model is named LSTM-Topic matrix factorization (LTMF) which integrates both LSTM and Topic Modeling for review understanding. In the experiments on popular review dataset Amazon , our LTMF model outperforms previous proposed HFT model and ConvMF model in rating prediction. Furthermore, LTMF shows the better ability on making topic clustering than traditional topic model based method, which implies integrating the information from deep learning and topic modeling is a meaningful approach to make a better understanding of reviews.

Community Member Retrieval on Social Media Using Textual Information

Aaron Jaech, Shobhit Hathi, and Mari Ostendorf

This paper addresses the problem of community membership detection using only text features in a scenario where a small number of positive labeled examples defines the community. The solution introduces an unsupervised proxy task for learning user embeddings: user re-identification. Experiments with 16 different communities show that the resulting embeddings are more effective for community membership identification than common unsupervised representations.

Cross-Domain Review Helpfulness Prediction Based on Convolutional Neural Networks with Auxiliary Domain Discriminators

Cen Chen, Yinfei Yang, Jun Zhou, Xiaolong Li, and Forrest Sheng Bao

With the growing amount of reviews in e-commerce websites, it is critical to assess the helpfulness of reviews and recommend them accordingly to consumers. Recent studies on review helpfulness require plenty of labeled samples for each domain/category of interests. However, such an approach based on close-world assumption is not always practical, especially for domains with limited reviews or the “out-of-vocabulary” problem. Therefore, we propose a convolutional neural network (CNN) based model which leverages both word-level and character-based representations. To transfer knowledge between domains, we further extend our model to jointly model different domains with auxiliary domain discriminators. On the Amazon product review dataset, our approach significantly outperforms the state of the art in terms of both accuracy and cross-domain robustness.

Deconfounded Lexicon Induction for Interpretable Social Science

Reid Pryzant, Kelly Shen, Dan Jurafsky, and Stefan Wagner

NLP algorithms are increasingly used in computational social science to take linguistic observations and predict outcomes like human preferences or actions. Making these social models transparent and interpretable often requires identifying features in the input that predict outcomes while also controlling for potential confounds. We formalize this need as a new task: inducing a lexicon that is predictive of a set of target variables yet uncorrelated to a set of confounding variables. We introduce two deep learning algorithms for the task. The first uses a bifurcated architecture to separate the explanatory power of the text and confounds. The second uses an adversarial discriminator to force confound-invariant text encodings. Both elicit lexicons from learned weights and attentional scores. We use them to induce lexicons that are predictive of timely responses to consumer complaints (controlling for product), enrollment from course descriptions (controlling for subject), and sales from product descriptions (controlling for seller). In each domain our algorithms pick words that are associated with *narrative persuasion*; more predictive and less confound-related than those of standard feature weighting and lexicon induction techniques like regression and log odds.

Detecting Denial-of-Service Attacks from Social Media Text: Applying NLP to Computer Security

Nathanael Chambers, Ben Fry, and James McMasters

This paper describes a novel application of NLP models to detect denial of service attacks using only social media as evidence. Individual networks are often slow in reporting attacks, so a detection system from public data could better assist a response to a broad attack across multiple services. We explore NLP methods to use social media as an indirect measure of network service status. We describe two learning frameworks for this task: a feed-forward neural network and a partially labeled LDA model. Both models outperform previous work by significant margins (20% F1 score). We further show that the topic-based model enables the first fine-grained analysis of how the public reacts to ongoing network attacks, discovering multiple “stages” of observation. This is the first model that both detects network attacks (with best performance) and provides an analysis of when and how the public interprets service outages. We describe the models, present experiments on the largest twitter DDoS corpus to date, and conclude with an analysis of public reactions based on the learned model’s output.

Predicting Foreign Language Usage from English-Only Social Media Posts

Svitlana Volkova, Stephen Ranshous, and Lawrence Phillips

Social media is known for its multi-cultural and multilingual interactions, a natural product of which is code-mixing. Multilingual speakers mix languages they tweet to address a different audience, express certain feelings, or attract attention. This paper presents a large-scale analysis of 6 million tweets produced by 27 thousand multilingual users speaking 12 other languages besides English. We rely on this corpus to build predictive models to infer non-English languages that users speak exclusively from their English tweets. Unlike native language identification task, we rely on large amounts of informal social media communications rather

than ESL essays. We contrast the predictive power of the state-of-the-art machine learning models trained on lexical, syntactic, and stylistic signals with neural network models learned from word, character and byte representations extracted from English only tweets. We report that content, style and syntax are the most predictive of non-English languages that users speak on Twitter. Neural network models learned from byte representations of user content combined with transfer learning yield the best performance. Finally, by analyzing cross-lingual transfer – the influence of non-English languages on various levels of linguistic performance in English, we present novel findings on stylistic and syntactic variations across speakers of 12 languages in social media.

The Importance of Calibration for Estimating Proportions from Annotations

Dallas Card and Noah A. Smith

Estimating label proportions in a target corpus is a type of measurement that is useful for answering certain types of social-scientific questions. While past work has described a number of relevant approaches, nearly all are based on an assumption which we argue is invalid for many problems, particularly when dealing with human annotations. In this paper, we identify and differentiate between two relevant data generating scenarios (intrinsic vs. extrinsic labels), introduce a simple but novel method which emphasizes the importance of calibration, and then analyze and experimentally validate the appropriateness of various methods for each of the two scenarios.

Corpus Creation and Emotion Prediction for Hindi-English Code-Mixed Social Media Text (SRW)

Deepanshu Vijay, Aditya Bohra, Vinay Singh, Syed Sarfaraz Akhtar, and Manish Shrivastava

Emotion Prediction is a Natural Language Processing (NLP) task dealing with detection and classification of emotions in various monolingual and bilingual texts. While some work has been done on code-mixed social media text and in emotion prediction separately, our work is the first attempt which aims at identifying the emotion associated with Hindi-English code-mixed social media text. In this paper, we analyze the problem of emotion identification in code-mixed content and present a Hindi-English code-mixed corpus extracted from twitter and annotated with the associated emotion. For every tweet in the dataset, we annotate the source language of all the words present, and also the causal language of the expressed emotion. Finally, we propose a supervised classification system which uses various machine learning techniques for detecting the emotion associated with the text using a variety of character level, word level, and lexicon based features.

Document-based Recommender System for Job Postings using Dense Representations (INDUSTRY)

Ahmed Elsafy, Martin Riedl, and Chris Biemann

Job boards and professional social networks heavily use recommender systems in order to better support users in exploring job advertisements. Detecting the similarity between job advertisements is important for job recommendation systems as it allows, for example, the application of item-to-item based recommendations. In this work, we research the usage of dense vector representations to enhance a large-scale job recommendation system and to rank German job advertisements regarding their similarity. We follow a two-folded evaluation scheme: (1) we exploit historic user interactions to automatically create a dataset of similar jobs that enables an offline evaluation. (2) In addition, we conduct an online A/B test and evaluate the best performing method on our platform reaching more than 1 million users. We achieve the best results by combining job titles with full-text job descriptions. In particular, this method builds dense document representation using words of the titles to weigh the importance of words of the full-text description. In the online evaluation, this approach allows us to increase the click-through rate on job recommendations for active users by 8.0%.

Posters: Summarization 2

Time: 10:30–12:00

Location: Elite Hall B

Chair: Lu Wang

A Dataset of Peer Reviews (PeerRead): Collection, Insights and NLP Applications

Dongyeop Kang, Waleed Ammar, Bhavana Dalvi, Madeleine van Zuylen, Sebastian Kohlmeier, Eduard Hovy, and Roy Schwartz

Peer reviewing is a central component in the scientific publishing process. We present the first public dataset of scientific peer reviews available for research purposes (PeerRead v1), providing an opportunity to study this important artifact. The dataset consists of 14.7K paper drafts and the corresponding accept/reject decisions in top-tier venues including ACL, NIPS and ICLR. The dataset also includes 10.7K textual peer reviews written

by experts for a subset of the papers. We describe the data collection process and report interesting observed phenomena in the peer reviews. We also propose two novel NLP tasks based on this dataset and provide simple baseline models. In the first task, we show that simple models can predict whether a paper is accepted with up to 21% error reduction compared to the majority baseline. In the second task, we predict the numerical scores of review aspects and show that simple models can outperform the mean baseline for aspects with high variance such as ‘originality’ and ‘impact’.

A Discourse-Aware Attention Model for Abstractive Summarization of Long Documents

Arman Cohan, Franck Dernoncourt, Doo Soon Kim, Trung Bui, Seokhwan Kim, Walter Chang, and Nazli Goharian

Neural abstractive summarization models have led to promising results in summarizing relatively short documents. We propose the first model for abstractive summarization of single, longer-form documents (e.g., research papers). Our approach consists of a new hierarchical encoder that models the discourse structure of a document, and an attentive discourse-aware decoder to generate the summary. Empirical results on two large-scale datasets of scientific papers show that our model significantly outperforms state-of-the-art models.

A Mixed Hierarchical Attention Based Encoder-Decoder Approach for Standard Table Summarization

Parag Jain, Anirban Laha, Karthik Sankaranarayanan, Preksha Nema, Mitesh M. Khapra, and Shreyas Shetty

Structured data summarization involves generation of natural language summaries from structured input data. In this work, we consider summarizing structured data occurring in the form of tables as they are prevalent across a wide variety of domains. We formulate the standard table summarization problem, which deals with tables conforming to a single predefined schema. To this end, we propose a mixed hierarchical attention based encoder-decoder model which is able to leverage the structure in addition to the content of the tables. Our experiments on the publicly available weathergov dataset show around 18 BLEU (around 30%) improvement over the current state-of-the-art.

Deep Communicating Agents for Abstractive Summarization

Asli Celikyilmaz, Antoine Bosselut, Xiaodong He, and Yejin Choi

We present deep communicating agents in an encoder-decoder architecture to address the challenges of representing a long document for abstractive summarization. With deep communicating agents, the task of encoding a long text is divided across multiple collaborating agents, each in charge of a subsection of the input text. These encoders are connected to a single decoder, trained end-to-end using reinforcement learning to generate a focused and coherent summary. Empirical results demonstrate that multiple communicating encoders lead to a higher quality summary compared to several strong baselines, including those based on a single encoder or multiple non-communicating encoders.

Effective Crowdsourcing for a New Type of Summarization Task

Youxuan Jiang, Catherine Finegan-Dollak, Jonathan K. Kummerfeld, and Walter Lasecki

Most summarization research focuses on summarizing the entire given text, but in practice readers are often interested in only one aspect of the document or conversation. We propose targeted summarization as an umbrella category for summarization tasks that intentionally consider only parts of the input data. This covers query-based summarization, update summarization, and a new task we propose where the goal is to summarize a particular aspect of a document. However, collecting data for this new task is hard because directly asking annotators (e.g., crowd workers) to write summaries leads to data with low accuracy when there are a large number of facts to include. We introduce a novel crowdsourcing workflow, Pin-Refine, that allows us to collect high-quality summaries for our task, a necessary step for the development of automatic systems.

Encoding Conversation Context for Neural Keyphrase Extraction from Microblog Posts

Yingyi Zhang, Jing Li, Yan Song, and Chengzhi Zhang

Existing keyphrase extraction methods suffer from data sparsity problem when they are conducted on short and informal texts, especially microblog messages. Enriching context is one way to alleviate this problem. Considering that conversations are formed by reposting and replying messages, they provide useful clues for recognizing essential content in target posts and are therefore helpful for keyphrase identification. In this paper, we present a neural keyphrase extraction framework for microblog posts that takes their conversation context into account, where four types of neural encoders, namely, averaged embedding, RNN, attention, and memory networks, are proposed to represent the conversation context. Experimental results on Twitter and Weibo datasets show that our framework with such encoders outperforms state-of-the-art approaches.

Estimating Summary Quality with Pairwise Preferences*Markus Zopf*

Automatic evaluation systems in the field of automatic summarization have been relying on the availability of gold standard summaries for over ten years. Gold standard summaries are expensive to obtain and often require the availability of domain experts to achieve high quality. In this paper, we propose an alternative evaluation approach based on pairwise preferences of sentences. In comparison to gold standard summaries, they are simpler and cheaper to obtain. In our experiments, we show that humans are able to provide useful feedback in the form of pairwise preferences. The new framework performs better than the three most popular versions of ROUGE with less expensive human input. We also show that our framework can reuse already available evaluation data and achieve even better results.

Generating Topic-Oriented Summaries Using Neural Attention*Kundan Krishna and Balaji Vasan Srinivasan*

Summarizing a document requires identifying the important parts of the document with an objective of providing a quick overview to a reader. However, a long article can span several topics and a single summary cannot do justice to all the topics. Further, the interests of readers can vary and the notion of importance can change across them. Existing summarization algorithms generate a single summary and are not capable of generating multiple summaries tuned to the interests of the readers. In this paper, we propose an attention based RNN framework to generate multiple summaries of a single document tuned to different topics of interest. Our method outperforms existing baselines and our results suggest that the attention of generative networks can be successfully biased to look at sentences relevant to a topic and effectively used to generate topic-tuned summaries.

Generative Bridging Network for Neural Sequence Prediction*Wenhu Chen, Guanlin Li, Shuo Ren, Shujie Liu, Zhirui Zhang, Mu Li, and Ming Zhou*

In order to alleviate data sparsity and overfitting problems in maximum likelihood estimation (MLE) for sequence prediction tasks, we propose the Generative Bridging Network (GBN), in which a novel bridge module is introduced to assist the training of the sequence prediction model (the generator network). Unlike MLE directly maximizing the conditional likelihood, the bridge extends the point-wise ground truth to a bridge distribution conditioned on it, and the generator is optimized to minimize their KL-divergence. Three different GBNs, namely uniform GBN, language-model GBN and coaching GBN, are proposed to penalize confidence, enhance language smoothness and relieve learning burden. Experiments conducted on two recognized sequence prediction tasks (machine translation and abstractive text summarization) show that our proposed GBNs can yield significant improvements over strong baselines. Furthermore, by analyzing samples drawn from different bridges, expected influences on the generator are verified.

Higher-order Syntactic Attention Network for Longer Sentence Compression*Hidetaka Kamigaito, Katsuhiro Hayashi, Tsutomu Hirao, and Masaaki Nagata*

A sentence compression method using LSTM can generate fluent compressed sentences. However, the performance of this method is significantly degraded when compressing longer sentences since it does not explicitly handle syntactic features. To solve this problem, we propose a higher-order syntactic attention network (HiSAN) that can handle higher-order dependency features as an attention distribution on LSTM hidden states. Furthermore, to avoid the influence of incorrect parse results, we trained HiSAN by maximizing jointly the probability of a correct output with the attention distribution. Experimental results on Google sentence compression dataset showed that our method achieved the best performance on F1 as well as ROUGE-1,2 and L scores, 83.2, 82.9, 75.8 and 82.7, respectively. In human evaluation, our methods also outperformed baseline methods in both readability and informativeness.

Key2Vec: Automatic Ranked Keyphrase Extraction from Scientific Articles using Phrase Embeddings*Debanjan Mahata, Rajiv Ratn Shah, John Kuriakose, and Roger Zimmermann*

Keyphrase extraction is a fundamental task in natural language processing that facilitates mapping of documents to a set of representative phrases. In this paper, we present an unsupervised technique (Key2Vec) that leverages phrase embeddings for ranking keyphrases extracted from scientific articles. Specifically, we propose an effective way of processing text documents for training multi-word phrase embeddings that are used for thematic representation of scientific articles and ranking of keyphrases extracted from them using theme-weighted PageRank. Evaluations are performed on benchmark datasets producing state-of-the-art results.

Learning to Generate Wikipedia Summaries for Underserved Languages from Wikidata
Lucie-Aimée Kaffee, Hady Elsahar, Pavlos Vougiouklis, Christophe Gravier, Frédérique Lafosse, Jonathon Hare, and Elena Simperl

While Wikipedia exists in 287 languages, its content is unevenly distributed among them. In this work, we investigate the generation of open domain Wikipedia summaries in underserved languages using structured data from Wikidata. To this end, we propose a neural network architecture equipped with copy actions that learns to generate single-sentence and comprehensible textual summaries from Wikidata triples. We demonstrate the effectiveness of the proposed approach by evaluating it against a set of baselines on two languages of different natures: Arabic, a morphological rich language with a larger vocabulary than English, and Esperanto, a constructed language known for its easy acquisition.

Multi-Reward Reinforced Summarization with Saliency and Entailment
Ramakanth Pasunuru and Mohit Bansal

Abstractive text summarization is the task of compressing and rewriting a long document into a short summary while maintaining saliency, directed logical entailment, and non-redundancy. In this work, we address these three important aspects of a good summary via a reinforcement learning approach with two novel reward functions: ROUGESal and Entail, on top of a coverage-based baseline. The ROUGESal reward modifies the ROUGE metric by up-weighting the salient phrases/words detected via a keyphrase classifier. The Entail reward gives high (length-normalized) scores to logically-entailed summaries using an entailment classifier. Further, we show superior performance improvement when these rewards are combined with traditional metric (ROUGE) based rewards, via our novel and effective multi-reward approach of optimizing multiple rewards simultaneously in alternate mini-batches. Our method achieves the new state-of-the-art results on CNN/Daily Mail dataset as well as strong improvements in a test-only transfer setup on DUC-2002.

Neural Storyline Extraction Model for Storyline Generation from News Articles
Deyu Zhou, Linsen Guo, and Yulan He

Storyline generation aims to extract events described on news articles under a certain news topic and reveal how those events evolve over time. Most approaches to storyline generation first train supervised models to extract events from news articles published in different time periods and then link relevant extracted events into coherent stories. They are domain dependent and cannot deal with unseen event types. To tackle this problem, approaches based on probabilistic graphic models jointly model the generations of events and storylines without the use of annotated data. However, the parameter inference procedure is too complex and models often require long time to converge. In this paper, we propose a novel neural network based approach to extract structured representations and evolution patterns of storylines without using annotated data. In this model, title and main body of a news article are assumed to share the similar storyline distribution. Moreover, similar documents described in neighboring time periods are assumed to share similar storyline distributions. Based on these assumptions, structured representations and evolution patterns of storylines can be extracted. The proposed model has been evaluated on three news corpora and the experimental results show that it outperforms state-of-the-art approaches for storyline generation on both accuracy and efficiency.

Objective Function Learning to Match Human Judgements for Optimization-Based Summarization
Maxime Peyrard and Iryna Gurevych

Supervised summarization systems usually rely on supervision at the sentence or n-gram level provided by automatic metrics like ROUGE, which act as noisy proxies for human judgments. In this work, we learn a summary-level scoring function θ including human judgments as supervision and automatically generated data as regularization. We extract summaries with a genetic algorithm using θ as a fitness function. We observe strong and promising performances across datasets in both automatic and manual evaluation.

Provable Fast Greedy Compressive Summarization with Any Monotone Submodular Function
Shinsaku Sakaue, Tsutomu Hirao, Masaaki Nishino, and Masaaki Nagata

Submodular maximization with the greedy algorithm has been studied as an effective approach to extractive summarization. This approach is known to have three advantages: its applicability to many useful submodular objective functions, the efficiency of the greedy algorithm, and the provable performance guarantee. However, when it comes to compressive summarization, we are currently missing a counterpart of the extractive method based on submodularity. In this paper, we propose a fast greedy method for compressive summarization. Our method is applicable to any monotone submodular objective function, including many functions well-suited for document summarization. We provide an approximation guarantee of our greedy algorithm. Experiments show

that our method is about 100 to 400 times faster than an existing method based on integer-linear-programming (ILP) formulations and that our method empirically achieves more than 95%-approximation.

Pruning Basic Elements for Better Automatic Evaluation of Summaries

Ukyo Honda, Tsutomu Hirao, and Masaaki Nagata

We propose a simple but highly effective automatic evaluation measure of summarization, pruned Basic Elements (pBE). Although the BE concept is widely used for the automated evaluation of summaries, its weakness is that it redundantly matches basic elements. To avoid this redundancy, pBE prunes basic elements by (1) disregarding frequency count of basic elements and (2) reducing semantically overlapped basic elements based on word similarity. Even though it is simple, pBE outperforms ROUGE in DUC datasets in most cases and achieves the highest rank correlation coefficient in TAC 2011 AESOP task.

Ranking Sentences for Extractive Summarization with Reinforcement Learning

Shashi Narayan, Shay B. Cohen, and Mirella Lapata

Single document summarization is the task of producing a shorter version of a document while preserving its principal information content. In this paper we conceptualize extractive summarization as a sentence ranking task and propose a novel training algorithm which globally optimizes the ROUGE evaluation metric through a reinforcement learning objective. We use our algorithm to train a neural summarization model on the CNN and DailyMail datasets and demonstrate experimentally that it outperforms state-of-the-art extractive and abstractive systems when evaluated automatically and by humans.

Relational Summarization for Corpus Analysis

Abram Handler and Brendan O'Connor

This work introduces a new problem, relational summarization, in which the goal is to generate a natural language summary of the relationship between two lexical items in a corpus, without reference to a knowledge base. Motivated by the needs of novel user interfaces, we define the task and give examples of its application. We also present a new query-focused method for finding natural language sentences which express relationships. Our method allows for summarization of more than two times more query pairs than baseline relation extractors, while returning measurably more readable output. Finally, to help guide future work, we analyze the challenges of relational summarization using both a news and a social media corpus.

Unsupervised Keyphrase Extraction with Multipartite Graphs

Florian Boudin

We propose an unsupervised keyphrase extraction model that encodes topical information within a multipartite graph structure. Our model represents keyphrase candidates and topics in a single graph and exploits their mutually reinforcing relationship to improve candidate ranking. We further introduce a novel mechanism to incorporate keyphrase selection preferences into the model. Experiments conducted on three widely used datasets show significant improvements over state-of-the-art graph-based models.

What's This Movie About? A Joint Neural Network Architecture for Movie Content Analysis

Philip John Gorinski and Mirella Lapata

This work takes a first step toward movie content analysis by tackling the novel task of movie overview generation. Overviews are natural language texts that give a first impression of a movie, describing aspects such as its genre, plot, mood, or artistic style. We create a dataset that consists of movie scripts, attribute-value pairs for the movies' aspects, as well as overviews, which we extract from an online database. We present a novel end-to-end model for overview generation, consisting of a multi-label encoder for identifying screenplay attributes, and an LSTM decoder to generate natural language sentences conditioned on the identified attributes. Automatic and human evaluation show that the encoder is able to reliably assign good labels for the movie's attributes, and the overviews provide descriptions of the movie's content which are informative and faithful.

Where Have I Heard This Story Before? Identifying Narrative Similarity in Movie Remakes

Snigdha Chaturvedi, Shashank Srivastava, and Dan Roth

People can identify correspondences between narratives in everyday life. For example, an analogy with the Cinderella story may be made in describing the unexpected success of an underdog in seemingly different stories. We present a new task and dataset for story understanding: identifying instances of similar narratives from a collection of narrative texts. We present an initial approach for this problem, which finds correspondences between narratives in terms of plot events, and resemblances between characters and their social relationships. Our approach yields an 8% absolute improvement in performance over a competitive information-retrieval

baseline on a novel dataset of plot summaries of 577 movie remakes from Wikipedia.

Which Scores to Predict in Sentence Regression for Text Summarization?

Markus Zopf, Eneldo Loza Mencía, and Johannes Fürnkranz

The task of automatic text summarization is to generate a short text that summarizes the most important information in a given set of documents. Sentence regression is an emerging branch in automatic text summarizations. Its key idea is to estimate the importance of information via learned utility scores for individual sentences. These scores are then used for selecting sentences from the source documents, typically according to a greedy selection strategy. Recently proposed state-of-the-art models learn to predict ROUGE recall scores of individual sentences, which seems reasonable since the final summaries are evaluated according to ROUGE recall. In this paper, we show in extensive experiments that following this intuition leads to suboptimal results and that learning to predict ROUGE precision scores leads to better results. The crucial difference is to aim not at covering as much information as possible but at wasting as little space as possible in every greedy step.

Demos

Time: 10:30–12:00

Location: Elite Hall B

Know Who Your Friends Are: Understanding Social Connections from Unstructured Text

Lea Deleris, Francesca Bonin, Elizabeth Daly, Stephane Deparis, Yufang Hou, Charles Jochim, Yassine Lassoued, and Killian Levacher

Having an understanding of interpersonal relationships is helpful in many contexts. Our system seeks to assist humans with that task, using textual information (e.g., case notes, speech transcripts, posts, books) as input. Specifically, our system first extracts qualitative and quantitative information elements (which we call signals) about interactions among persons, aggregates those to provide a condensed view of relationships and then enables users to explore all facets of the resulting social (multi-)graph through a visual interface.

RiskFinder: A Sentence-level Risk Detector for Financial Reports

Yu-Wen Liu, Liang-Chih Liu, Chuan-Ju Wang, and Ming-Feng Tsai

This paper presents a web-based information system, RiskFinder, for facilitating the analyses of soft and hard information in financial reports. In particular, the system broadens the analyses from the word level to sentence level, which makes the system useful for practitioner communities and unprecedented among financial academics. The proposed system has four main components: 1) a Form 10-K risk-sentiment dataset, consisting of a set of risk-labeled financial sentences and pre-trained sentence embeddings; 2) metadata, including basic information on each company that published the Form 10-K financial report as well as several relevant financial measures; 3) an interface that highlights risk-related sentences in the financial reports based on the latest sentence embedding techniques; 4) a visualization of financial time-series data for a corresponding company. This paper also conducts some case studies to showcase that the system can be of great help in capturing valuable insight within large amounts of textual information. The system is now online available at <https://cfda.csie.org/RiskFinder/>.

Session 11 Overview – Monday, June 4, 2018

	Track A <i>Sentiment Analysis 2</i> Empire A	Track B <i>Discourse and Pragmatics 2</i> Empire B	Track C <i>Tagging, Chunking, Syntax and Parsing 3</i> Empire C	Track P <i>Posters and Demos</i> Elite Hall B
14:00	Sentiment Analysis: It's Complicated! <i>K. Kenyon-Dean, E. Ahmed, S. Fujimoto, J. Georges-Flitteau, C. Glaz, B. Kaur, A. Lalande, S. Bhandari, R. Belfer, N. Kanagasabai, R. Sarrazingendron, R. Verma, and D. Ruths</i>	The Argument Reasoning Comprehension Task: Identification and Reconstruction of Implicit Warrants <i>I. Habernal, H. Wachsmuth, I. Gurevych, and B. Stein</i>	Do Latent Tree Learning Models Identify Meaningful Structure in Sentences? (TACL) <i>A. Williams, A. Drozdov, and S. Bowman</i>	
14:18	Multi-Task Learning of Pairwise Sequence Classification Tasks over Disparate Label Spaces <i>I. Augenstein, S. Ruder, and Å. Søgaard</i>	Linguistic Cues to Deception and Perceived Deception in Interview Dialogues <i>S. I. Levitan, A. Maredia, and J. Hirschberg</i>	Unsupervised Acquisition of Comprehensive Multiword Lexicons using Competition in an N-Gram Lattice (TACL) <i>J. Brooke, J. Snajder, and T. Baldwin</i>	
14:36	Word Emotion Induction for Multiple Languages as a Deep Multi-Task Learning Problem <i>S. Buechel and U. Hahn</i>	Unified Pragmatic Models for Generating and Following Instructions <i>D. Fried, J. Andreas, and D. Klein</i>	Unsupervised Grammar Induction with Depth-bounded PCFG (TACL) <i>L. Jin, F. Doshi-Velez, T. Miller, W. Schuler, and L. Schwartz</i>	Cognitive Modeling and Psycholinguistics 2 / Dialogue and Interactive Systems 2 / Text Mining 2 / Speech 2 / Vision, Robotics and Other Grounding 3 / Demos
14:54	Human Needs Categorization of Affective Events Using Labeled and Unlabeled Data <i>H. Ding and E. Riloff</i>	Hierarchical Structured Model for Fine-to-Coarse Manifesto Text Analysis <i>S. Subramanian, T. Cohn, and T. Baldwin</i>	Non-Projective Dependency Parsing with Non-Local Transitions <i>D. Fernández-González and C. Gómez-Rodríguez</i>	
15:12	Multimodal Emoji Prediction <i>F. Barbieri, M. Ballessteros, F. Ronzano, and H. Saggion</i>	Higher-order Coreference Resolution with Coarse-to-Fine Inference <i>K. Lee, L. He, and L. Zettlemoyer</i>	Behavior Analysis of NLI Models: Uncovering the Influence of Three Factors on Robustness <i>I. Sanchez, J. Mitchell, and S. Riedel</i>	

Parallel Session 11

Oral: Sentiment Analysis 2

Empire A

Chair: Smaranda Muresan

Sentiment Analysis: It's Complicated!

Kian Kenyon-Dean, Eisha Ahmed, Scott Fujimoto, Jeremy Georges-Filteau, Christopher Glasz, Barleen Kaur, Auguste Lalande, Shruti Bhanderi, Robert Belfer, Nirmal Kanagasabai, Roman Sarrazingendron, Rohit Verma, and Derek Ruths

14:00–14:17

Sentiment analysis is used as a proxy to measure human emotion, where the objective is to categorize text according to some predefined notion of sentiment. Sentiment analysis datasets are typically constructed with gold-standard sentiment labels, assigned based on the results of manual annotations. When working with such annotations, it is common for dataset constructors to discard “noisy” or “controversial” data where there is significant disagreement on the proper label. In datasets constructed for the purpose of Twitter sentiment analysis (TSA), these controversial examples can compose over 30% of the originally annotated data. We argue that the removal of such data is a problematic trend because, when performing real-time sentiment classification of short-text, an automated system cannot know *a priori* which samples would fall into this category of disputed sentiment. We therefore propose the notion of a “complicated” class of sentiment to categorize such text, and argue that its inclusion in the short-text sentiment analysis framework will improve the quality of automated sentiment analysis systems as they are implemented in real-world settings. We motivate this argument by building and analyzing a new publicly available TSA dataset of over 7,000 tweets annotated with 5x coverage, named MTS. Our analysis of classifier performance over our dataset offers insights into sentiment analysis dataset and model design, how current techniques would perform in the real world, and how researchers should handle difficult data.

Multi-Task Learning of Pairwise Sequence Classification Tasks over Disparate Label Spaces

Isabelle Augenstein, Sebastian Ruder, and Anders Søgaard

14:18–14:35

We combine multi-task learning and semi-supervised learning by inducing a joint embedding space between disparate label spaces and learning transfer functions between label embeddings, enabling us to jointly leverage unlabelled data and auxiliary, annotated datasets. We evaluate our approach on a variety of tasks with disparate label spaces. We outperform strong single and multi-task baselines and achieve a new state of the art for aspect-based and topic-based sentiment analysis.

Word Emotion Induction for Multiple Languages as a Deep Multi-Task Learning Problem

Sven Buechel and Udo Hahn

14:36–14:53

Predicting the emotional value of lexical items is a well-known problem in sentiment analysis. While research has focused on polarity for quite a long time, meanwhile this early focus has been shifted to more expressive emotion representation models (such as Basic Emotions or Valence-Arousal-Dominance). This change resulted in a proliferation of heterogeneous formats and, in parallel, often small-sized, non-interoperable resources (lexicons and corpus annotations). In particular, the limitations in size hampered the application of deep learning methods in this area because they typically require large amounts of input data. We here present a solution to get around this language data bottleneck by rephrasing word emotion induction as a multi-task learning problem. In this approach, the prediction of each independent emotion dimension is considered as an individual task and hidden layers are shared between these dimensions. We investigate whether multi-task learning is more advantageous than single-task learning for emotion prediction by comparing our model against a wide range of alternative emotion and polarity induction methods featuring 9 typologically diverse languages and a total of 15 conditions. Our model turns out to outperform each one of them. Against all odds, the proposed deep learning approach yields the largest gain on the smallest data sets, merely composed of one thousand samples.

Human Needs Categorization of Affective Events Using Labeled and Unlabeled Data

Haibo Ding and Ellen Riloff

14:54–15:12

We often talk about events that impact us positively or negatively. For example “I got a job” is good news, but “I lost my job” is bad news. When we discuss an event, we not only understand its affective polarity but also the reason why the event is beneficial or detrimental. For example, getting or losing a job has affective polarity primarily because it impacts us financially. Our work aims to categorize affective events based upon

human need categories that often explain people's motivations and desires: PHYSIOLOGICAL, HEALTH, LEISURE, SOCIAL, FINANCIAL, COGNITION, and FREEDOM. We create classification models based on event expressions as well as models that use contexts surrounding event mentions. We also design a co-training model that learns from unlabeled data by simultaneously training event expression and event context classifiers in an iterative learning process. Our results show that co-training performs well, producing substantially better results than the individual classifiers.

Multimodal Emoji Prediction

Francesco Barbieri, Miguel Ballesteros, Francesco Ronzano, and Horacio Saggion 15:13–15:30

Emojis are small images that are commonly included in social media text messages. The combination of visual and textual content in the same message builds up a modern way of communication, that automatic systems are not used to deal with. In this paper we extend recent advances in emoji prediction by putting forward a multimodal approach that is able to predict emojis in Instagram posts. Instagram posts are composed of pictures together with texts which sometimes include emojis. We show that these emojis can be predicted by using the text, but also using the picture. Our main finding is that incorporating the two synergistic modalities, in a combined model, improves accuracy in an emoji prediction task. This result demonstrates that these two modalities (text and images) encode different information on the use of emojis and therefore can complement each other.

Oral: Discourse and Pragmatics 2

Empire B

Chair: Vincent Ng

The Argument Reasoning Comprehension Task: Identification and Reconstruction of Implicit Warrants

Ivan Habernal, Henning Wachsmuth, Iryna Gurevych, and Benno Stein

14:00–14:17

Reasoning is a crucial part of natural language argumentation. To comprehend an argument, one must analyze its warrant, which explains why its claim follows from its premises. As arguments are highly contextualized, warrants are usually presupposed and left implicit. Thus, the comprehension does not only require language understanding and logic skills, but also depends on common sense. In this paper we develop a methodology for reconstructing warrants systematically. We operationalize it in a scalable crowdsourcing process, resulting in a freely licensed dataset with warrants for 2k authentic arguments from news comments. On this basis, we present a new challenging task, the argument reasoning comprehension task. Given an argument with a claim and a premise, the goal is to choose the correct implicit warrant from two options. Both warrants are plausible and lexically close, but lead to contradicting claims. A solution to this task will define a substantial step towards automatic warrant reconstruction. However, experiments with several neural attention and language models reveal that current approaches do not suffice.

Linguistic Cues to Deception and Perceived Deception in Interview Dialogues

Sarah Ita Levitan, Angel Maredia, and Julia Hirschberg

14:18–14:35

We explore deception detection in interview dialogues. We analyze a set of linguistic features in both truthful and deceptive responses to interview questions. We also study the perception of deception, identifying characteristics of statements that are perceived as truthful or deceptive by interviewers. Our analysis show significant differences between truthful and deceptive question responses, as well as variations in deception patterns across gender and native language. This analysis motivated our selection of features for machine learning experiments aimed at classifying globally deceptive speech. Our best classification performance is 72.74% F1-Score (about 17% better than human performance), which is achieved using a combination of linguistic features and individual traits.

Unified Pragmatic Models for Generating and Following Instructions

Daniel Fried, Jacob Andreas, and Dan Klein

14:36–14:53

We show that explicit pragmatic inference aids in correctly generating and following natural language instructions for complex, sequential tasks. Our pragmatics-enabled models reason about why speakers produce certain instructions, and about how listeners will react upon hearing them. Like previous pragmatic models, we use learned base listener and speaker models to build a pragmatic speaker that uses the base listener to simulate the interpretation of candidate descriptions, and a pragmatic listener that reasons counterfactually about alternative descriptions. We extend these models to tasks with sequential structure. Evaluation of language generation and interpretation shows that pragmatic inference improves state-of-the-art listener models (at cor-

rectly interpreting human instructions) and speaker models (at producing instructions correctly interpreted by humans) in diverse settings.

Hierarchical Structured Model for Fine-to-Coarse Manifesto Text Analysis

Shivashankar Subramanian, Trevor Cohn, and Timothy Baldwin

14:54–15:12

Election manifestos document the intentions, motives, and views of political parties. They are often used for analysing a party's fine-grained position on a particular issue, as well as for coarse-grained positioning of a party on the left-right spectrum. In this paper we propose a two-stage model for automatically performing both levels of analysis over manifestos. In the first step we employ a hierarchical multi-task structured deep model to predict fine- and coarse-grained positions, and in the second step we perform post-hoc calibration of coarse-grained positions using probabilistic soft logic. We empirically show that the proposed model outperforms state-of-art approaches at both granularities using manifestos from twelve countries, written in ten different languages.

Higher-order Coreference Resolution with Coarse-to-Fine Inference

Kenton Lee, Luheng He, and Luke Zettlemoyer

15:13–15:30

We introduce a fully-differentiable approximation to higher-order inference for coreference resolution. Our approach uses the antecedent distribution from a span-ranking architecture as an attention mechanism to iteratively refine span representations. This enables the model to softly consider multiple hops in the predicted clusters. To alleviate the computational cost of this iterative process, we introduce a coarse-to-fine approach that incorporates a less accurate but more efficient bilinear factor, enabling more aggressive pruning without hurting accuracy. Compared to the existing state-of-the-art span-ranking approach, our model significantly improves accuracy on the English OntoNotes benchmark, while being far more computationally efficient.

Oral: Tagging, Chunking, Syntax and Parsing 3

Empire C

Chair: Thamar Solorio

Do Latent Tree Learning Models Identify Meaningful Structure in Sentences? (TACL)

Adina Williams, Andrew Drozdov, and Samuel Bowman

14:00–14:17

Recent work on the problem of latent tree learning has made it possible to train neural networks that learn to both parse a sentence and use the resulting parse to interpret the sentence, all without exposure to ground-truth parse trees at training time. Surprisingly, these models often perform better at sentence understanding tasks than models that use parse trees from conventional parsers. This paper aims to investigate what these latent tree learning models learn. We replicate two such models in a shared codebase and find that (i) only one of these models outperforms conventional tree-structured models on sentence classification, (ii) its parsing strategies are not especially consistent across random restarts, (iii) the parses it produces tend to be shallower than standard Penn Treebank (PTB) parses, and (iv) they do not resemble those of PTB or any other semantic or syntactic formalism that the authors are aware of.

Unsupervised Acquisition of Comprehensive Multiword Lexicons using Competition in an N-Gram Lattice (TACL)

Julian Brooke, Jan Šnajder, and Timothy Baldwin

14:18–14:35

We present a new model for acquiring comprehensive multiword lexicons from large corpora based on competition among n-gram candidates. In contrast to the standard approach of simple ranking by association measure, in our model n-grams are arranged in a lattice structure based on subsumption and overlap relationships, with nodes inhibiting other nodes in their vicinity when they are selected as a lexical item. We show how the configuration of such a lattice can be optimized tractably, and demonstrate using annotations of sampled n-grams that our method consistently outperforms alternatives by at least 0.05 F-score across several corpora and languages.

Unsupervised Grammar Induction with Depth-bounded PCFG (TACL)

Lifeng Jin, Finale Doshi-Velez, Timothy Miller, William Schuler, and Lane Schwartz 14:36–14:53

There has been recent interest in applying cognitively or empirically motivated bounds on recursion depth to limit the search space of grammar induction models (Ponvert et al., 2011; Noji and Johnson, 2016; Shain et al., 2016). This work extends this depth-bounding approach to probabilistic context-free grammar induction (DB-PCFG), which has a smaller parameter space than hierarchical sequence models, and therefore more fully exploits the space reductions of depth-bounding. Results for this model on grammar acquisition from transcribed child-directed speech and newswire text exceed or are competitive with those of other models

when evaluated on parse accuracy. Moreover, grammars acquired from this model demonstrate a consistent use of category labels, something which has not been demonstrated by other acquisition models.

Non-Projective Dependency Parsing with Non-Local Transitions

Daniel Fernández-González and Carlos Gómez-Rodríguez

14:54–15:12

We present a novel transition system, based on the Covington non-projective parser, introducing non-local transitions that can directly create arcs involving nodes to the left of the current focus positions. This avoids the need for long sequences of No-Arcs transitions to create long-distance arcs, thus alleviating error propagation. The resulting parser outperforms the original version and achieves the best accuracy on the Stanford Dependencies conversion of the Penn Treebank among greedy transition-based parsers.

Behavior Analysis of NLI Models: Uncovering the Influence of Three Factors on Robustness

Ivan Sanchez, Jeff Mitchell, and Sebastian Riedel

15:13–15:30

Natural Language Inference is a challenging task that has received substantial attention, and state-of-the-art models now achieve impressive test set performance in the form of accuracy scores. Here, we go beyond this single evaluation metric to examine robustness to semantically-valid alterations to the input data. We identify three factors - insensitivity, polarity and unseen pairs - and compare their impact on three SNLI models under a variety of conditions. Our results demonstrate a number of strengths and weaknesses in the models' ability to generalise to new in-domain instances. In particular, while strong performance is possible on unseen hypernyms, unseen antonyms are more challenging for all the models. More generally, the models suffer from an insensitivity to certain small but semantically significant alterations, and are also often influenced by simple statistical correlations between words and training labels. Overall, we show that evaluations of NLI models can benefit from studying the influence of factors intrinsic to the models or found in the dataset used.

Posters and Demos: Session 11

Posters: Cognitive Modeling and Psycholinguistics 2

Time: 14:00–15:30

Location: Elite Hall B

Chair: Hannaneh Hajishirzi

Assessing Language Proficiency from Eye Movements in Reading

Yevgeni Berzak, Boris Katz, and Roger Levy

We present a novel approach for determining learners' second language proficiency which utilizes behavioral traces of eye movements during reading. Our approach provides stand-alone eyetracking based English proficiency scores which reflect the extent to which the learner's gaze patterns in reading are similar to those of native English speakers. We show that our scores correlate strongly with standardized English proficiency tests. We also demonstrate that gaze information can be used to accurately predict the outcomes of such tests. Our approach yields the strongest performance when the test taker is presented with a suite of sentences for which we have eyetracking data from other readers. However, it remains effective even using eyetracking with sentences for which eye movement data have not been previously collected. By deriving proficiency as an automatic byproduct of eye movements during ordinary reading, our approach offers a potentially valuable new tool for second language proficiency assessment. More broadly, our results open the door to future methods for inferring reader characteristics from the behavioral traces of reading.

Comparing Theories of Speaker Choice Using a Model of Classifier Production in Mandarin Chinese

Meilin Zhan and Roger Levy

Speakers often have more than one way to express the same meaning. What general principles govern speaker choice in the face of optionality when near semantically invariant alternation exists? Studies have shown that optional reduction in language is sensitive to contextual predictability, such that more predictable a linguistic unit is, the more likely it is to get reduced. Yet it is unclear whether these cases of speaker choice are driven by audience design versus toward facilitating production. Here we argue that for a different optionality phenomenon, namely classifier choice in Mandarin Chinese, Uniform Information Density and at least one plausible variant of availability-based production make opposite predictions regarding the relationship between the predictability of the upcoming material and speaker choices. In a corpus analysis of Mandarin Chinese, we show that the distribution of speaker choices supports the availability-based production account and not the Uniform Information Density.

Detecting Linguistic Characteristics of Alzheimer's Dementia by Interpreting Neural Models

Sweta Karlekar, Tong Niu, and Mohit Bansal

Alzheimer's disease (AD) is an irreversible and progressive brain disease that can be stopped or slowed down with medical treatment. Language changes serve as a sign that a patient's cognitive functions have been impacted, potentially leading to early diagnosis. In this work, we use NLP techniques to classify and analyze the linguistic characteristics of AD patients using the DementiaBank dataset. We apply three neural models based on CNNs, LSTM-RNNs, and their combination, to distinguish between language samples from AD and control patients. We achieve a new independent benchmark accuracy for the AD classification task. More importantly, we next interpret what these neural models have learned about the linguistic characteristics of AD patients, via analysis based on activation clustering and first-derivative saliency techniques. We then perform novel automatic pattern discovery inside activation clusters, and consolidate AD patients' distinctive grammar patterns. Additionally, we show that first derivative saliency can not only rediscover previous language patterns of AD patients, but also shed light on the limitations of neural models. Lastly, we also include analysis of gender-separated AD data.

Spotting Spurious Data with Neural Networks

Hadi Amiri, Timothy Miller, and Guergana Savova

Automatic identification of spurious instances (those with potentially wrong labels in datasets) can improve the quality of existing language resources, especially when annotations are obtained through crowdsourcing or automatically generated based on coded rankings. In this paper, we present effective approaches inspired by queueing theory and psychology of learning to automatically identify spurious instances in datasets. Our approaches discriminate instances based on their "difficulty to learn," determined by a downstream learner.

Our methods can be applied to any dataset assuming the existence of a neural network model for the target task of the dataset. Our best approach outperforms competing state-of-the-art baselines and has a MAP of 0.85 and 0.22 in identifying spurious instances in synthetic and carefully-crowdsourced real-world datasets respectively.

The Timing of Lexical Memory Retrievals in Language Production*Jeremy Cole and David Reitter*

This paper explores the time course of lexical memory retrieval by modeling fluent language production. The duration of retrievals is predicted using the ACT-R cognitive architecture. In a large-scale observational study of a spoken corpus, we find that language production at a time point preceding a word is sped up or slowed down depending on activation of that word. This computational analysis has consequences for the theoretical model of language production. The results point to interference between lexical and phonological stages as well as a quantifiable buffer for lexical information that opens up the possibility of non-sequential retrievals.

Unsupervised Induction of Linguistic Categories with Records of Reading, Speaking, and Writing*Maria Barrett, Ana Valeria Gonzalez-Garduño, Lea Frermann, and Anders Søgaard*

When learning POS taggers and syntactic chunkers for low-resource languages, different resources may be available, and often all we have is a small tag dictionary, motivating type-constrained unsupervised induction. Even small dictionaries can improve the performance of unsupervised induction algorithms. This paper shows that performance can be further improved by including data that is readily available or can be easily obtained for most languages, i.e., eye-tracking, speech, or keystroke logs (or any combination thereof). We project information from all these data sources into shared spaces, in which the union of words is represented. For English unsupervised POS induction, the additional information, which is not required at test time, leads to an average error reduction on Ontonotes domains of 1.5% over systems augmented with state-of-the-art word embeddings. On Penn Treebank the best model achieves 5.4% error reduction over a word embeddings baseline. We also achieve significant improvements for syntactic chunk induction. Our analysis shows that improvements are even bigger when the available tag dictionaries are smaller.

Sensing and Learning Human Annotators Engaged in Narrative Sensemaking (SRW)*McKenna Tornblad, Luke Lapresi, Christopher Homan, Raymond Ptucha, and Cecilia Ovesdotter Alm*

While labor issues and quality assurance in crowdwork are increasingly studied, how annotators make sense of texts and how they are personally impacted by doing so are not. We study these questions via a narrative-sorting annotation task, where carefully selected (by sequentiality, topic, emotional content, and length) collections of tweets serve as examples of everyday storytelling. As readers process these narratives, we measure their facial expressions, galvanic skin response, and self-reported reactions. From the perspective of annotator well-being, a reassuring outcome was that the sorting task did not cause a measurable stress response, however readers reacted to humor. In terms of sensemaking, readers were more confident when sorting sequential, target-topical, and highly emotional tweets. As crowdsourcing becomes more common, this research sheds light onto the perceptive capabilities and emotional impact of human readers.

Posters: Dialogue and Interactive Systems 2

Time: 14:00–15:30

Location: Elite Hall B

Chair: Hannaneh Hajishirzi

Challenging Reading Comprehension on Daily Conversation: Passage Completion on Multiparty Dialog*Kaixin Ma, Tomasz Jurczyk, and Jinho D. Choi*

This paper presents a new corpus and a robust deep learning architecture for a task in reading comprehension, passage completion, on multiparty dialog. Given a dialog in text and a passage containing factual descriptions about the dialog where mentions of the characters are replaced by blanks, the task is to fill the blanks with the most appropriate character names that reflect the contexts in the dialog. Since there is no dataset that challenges the task of passage completion in this genre, we create a corpus by selecting transcripts from a TV show that comprise 1,681 dialogs, generating passages for each dialog through crowdsourcing, and annotating mentions of characters in both the dialog and the passages. Given this dataset, we build a deep neural model that integrates rich feature extraction from convolutional neural networks into sequence modeling in recurrent neural networks, optimized by utterance and dialog level attentions. Our model outperforms the previous state-

of-the-art model on this task in a different genre using bidirectional LSTM, showing a 13.0+% improvement for longer dialogs. Our analysis shows the effectiveness of the attention mechanisms and suggests a direction to machine comprehension on multiparty dialog.

Deep Dungeons and Dragons: Learning Character-Action Interactions from Role-Playing Game Transcripts

Annie Louis and Charles Sutton

An essential aspect to understanding narratives is to grasp the interaction between characters in a story and the actions they take. We examine whether computational models can capture this interaction, when both character attributes and actions are expressed as complex natural language descriptions. We propose role-playing games as a testbed for this problem, and introduce a large corpus of game transcripts collected from online discussion forums. Using neural language models which combine character and action descriptions from these stories, we show that we can learn the latent ties. Action sequences are better predicted when the character performing the action is also taken into account, and vice versa for character attributes.

Dialog Generation Using Multi-Turn Reasoning Neural Networks

Xianchao Wu, Ander Martinez, and Momo Klyen

In this paper, we propose a generalizable dialog generation approach that adapts multi-turn reasoning, one recent advancement in the field of document comprehension, to generate responses ("answers") by taking current conversation session context as a "document" and current query as a "question". The major idea is to represent a conversation session into memories upon which attention-based memory reading mechanism can be performed multiple times, so that (1) user's query is properly extended by contextual clues and (2) optimal responses are step-by-step generated. Considering that the speakers of one conversation are not limited to be one, we separate the single memory used for document comprehension into different groups for speaker-specific topic and opinion embedding. Namely, we utilize the queries' memory, the responses' memory, and their unified memory, following the time sequence of the conversation session. Experiments on Japanese 10-sentence (5-round) conversation modeling show impressive results on how multi-turn reasoning can produce more diverse and acceptable responses than state-of-the-art single-turn and non-reasoning baselines.

Dialogue Learning with Human Teaching and Feedback in End-to-End Trainable Task-Oriented Dialogue Systems

Bing Liu, Gokhan Tür, Dilek Hakkani-Tür, Pararth Shah, and Larry Heck

In this work, we present a hybrid learning method for training task-oriented dialogue systems through online user interactions. Popular methods for learning task-oriented dialogues include applying reinforcement learning with user feedback on supervised pre-training models. Efficiency of such learning method may suffer from the mismatch of dialogue state distribution between offline training and online interactive learning stages. To address this challenge, we propose a hybrid imitation and reinforcement learning method, with which a dialogue agent can effectively learn from its interaction with users by learning from human teaching and feedback. We design a neural network based task-oriented dialogue agent that can be optimized end-to-end with the proposed learning method. Experimental results show that our end-to-end dialogue agent can learn effectively from the mistake it makes via imitation learning from user teaching. Applying reinforcement learning with user feedback after the imitation learning stage further improves the agent's capability in successfully completing a task.

Feudal Reinforcement Learning for Dialogue Management in Large Domains

Iñigo Casanueva, Paweł Budzianowski, Pei-Hao Su, Stefan Ultes, Lina M. Rojas Barahona, Bo-Hsiang Tseng, and Milica Gasic

Reinforcement learning (RL) is a promising approach to solve dialogue policy optimisation. Traditional RL algorithms, however, fail to scale to large domains due to the curse of dimensionality. We propose a novel Dialogue Management architecture, based on Feudal RL, which decomposes the decision into two steps; a first step where a master policy selects a subset of primitive actions, and a second step where a primitive action is chosen from the selected subset. The structural information included in the domain ontology is used to abstract the dialogue state space, taking the decisions at each step using different parts of the abstracted state. This, combined with an information sharing mechanism between slots, increases the scalability to large domains. We show that an implementation of this approach, based on Deep-Q Networks, significantly outperforms previous state of the art in several dialogue domains and environments, without the need of any additional reward signal.

LSDSCC: a Large Scale Domain-Specific Conversational Corpus for Response Generation with Diversity Oriented Evaluation Metrics

Zhen Xu, Nan Jiang, Bingquan Liu, Wenge Rong, Bowen Wu, Baoxun Wang, Zhuoran Wang, and Xiaolong Wang

It has been proven that automatic conversational agents can be built up using the End-to-End Neural Response Generation (NRG) framework, and such a data-driven methodology requires a large number of dialog pairs for model training and reasonable evaluation metrics for testing. This paper proposes a Large Scale Domain-Specific Conversational Corpus (LSDSCC) composed of high-quality query-response pairs extracted from the domain-specific online forum, with thorough preprocessing and cleansing procedures. Also, a testing set, including multiple diverse responses annotated for each query, is constructed, and on this basis, the metrics for measuring the diversity of generated results are further presented. We evaluate the performances of neural dialog models with the widely applied diversity boosting strategies on the proposed dataset. The experimental results have shown that our proposed corpus can be taken as a new benchmark dataset for the NRG task, and the presented metrics are promising to guide the optimization of NRG models by quantifying the diversity of the generated responses reasonably.

Posters: Text Mining 2

Time: 14:00–15:30

Location: Elite Hall B

Chair: Hannaneh Hajishirzi

EMR Coding with Semi-Parametric Multi-Head Matching Networks

Anthony Rios and Ramakanth Kavuluru

Coding EMRs with diagnosis and procedure codes is an indispensable task for billing, secondary data analyses, and monitoring health trends. Both speed and accuracy of coding are critical. While coding errors could lead to more patient-side financial burden and misinterpretation of a patient's well-being, timely coding is also needed to avoid backlog and additional costs for the healthcare facility. In this paper, we present a new neural network architecture that combines ideas from few-shot learning matching networks, multi-label loss functions, and convolutional neural networks for text classification to significantly outperform other state-of-the-art models. Our evaluations are conducted using a well known de-identified EMR dataset (MIMIC) with a variety of multi-label performance measures.

Evaluating Historical Text Normalization Systems: How Well Do They Generalize?

Alexander Robertson and Sharon Goldwater

We highlight several issues in the evaluation of historical text normalization systems that make it hard to tell how well these systems would actually work in practice—i.e., for new datasets or languages; in comparison to more naïve systems; or as a preprocessing step for downstream NLP tools. We illustrate these issues and exemplify our proposed evaluation practices by comparing two neural models against a naïve baseline system. We show that the neural models generalize well to unseen words in tests on five languages; nevertheless, they provide no clear benefit over the naïve baseline for downstream POS tagging of an English historical collection. We conclude that future work should include more rigorous evaluation, including both intrinsic and extrinsic measures where possible.

Gated Multi-Task Network for Text Classification

Liqiang Xiao, Honglun Zhang, and Wenqing Chen

Multi-task learning with Convolutional Neural Network (CNN) has shown great success in many Natural Language Processing (NLP) tasks. This success can be largely attributed to the feature sharing by fusing some layers among tasks. However, most existing approaches just fully or proportionally share the features without distinguishing the helpfulness of them. By that the network would be confused by the helpless even harmful features, generating undesired interference between tasks. In this paper, we introduce gate mechanism into multi-task CNN and propose a new Gated Sharing Unit, which can filter the feature flows between tasks and greatly reduce the interference. Experiments on 9 text classification datasets shows that our approach can learn selection rules automatically and gain a great improvement over strong baselines.

Mining Evidences for Concept Stock Recommendation

Qi Liu and Yue Zhang

We investigate the task of mining relevant stocks given a topic of concern on emerging capital markets, for which there is lack of structural understanding. Deep learning is leveraged to mine evidences from large scale textual data, which contain valuable market information. In particular, distributed word similarities trained

over large scale raw texts are taken as a basis of relevance measuring, and deep reinforcement learning is leveraged to learn a strategy of topic expansion, given a small amount of manually labeled data from financial analysts. Results on two Chinese stock market datasets show that our method outperforms a strong baseline using information retrieval techniques.

Factors Influencing the Surprising Instability of Word Embeddings

Laura Wendlandt, Jonathan K. Kummerfeld, and Rada Mihalcea

Despite the recent popularity of word embedding methods, there is only a small body of work exploring the limitations of these representations. In this paper, we consider one aspect of embedding spaces, namely their stability. We show that even relatively high frequency words (100-200 occurrences) are often unstable. We provide empirical evidence for how various factors contribute to the stability of word embeddings, and we analyze the effects of stability on downstream tasks.

Natural Language to Structured Query Generation via Meta-Learning

Po-Sen Huang, Chenglong Wang, Rishabh Singh, Wen-tau Yih, and Xiaodong He

In conventional supervised training, a model is trained to fit all the training examples. However, having a monolithic model may not always be the best strategy, as examples could vary widely. In this work, we explore a different learning protocol that treats each example as a unique pseudo-task, by reducing the original learning problem to a few-shot meta-learning scenario with the help of a domain-dependent relevance function. When evaluated on the WikiSQL dataset, our approach leads to faster convergence and achieves 1.1%–5.4% absolute accuracy gains over the non-meta-learning counterparts.

Smaller Text Classifiers with Discriminative Cluster Embeddings

Mingda Chen and Kevin Gimpel

Word embedding parameters often dominate overall model sizes in neural methods for natural language processing. We reduce deployed model sizes of text classifiers by learning a hard word clustering in an end-to-end manner. We use the Gumbel-Softmax distribution to maximize over the latent clustering while minimizing the task loss. We propose variations that selectively assign additional parameters to words, which further improves accuracy while still remaining parameter-efficient.

Posters: Speech 2

Time: 14:00–15:30

Location: Elite Hall B

Chair: Hannaneh Hajishirzi

Binarized LSTM Language Model

Xuan Liu, Di Cao, and Kai Yu

Long short-term memory (LSTM) language model (LM) has been widely investigated for automatic speech recognition (ASR) and natural language processing (NLP). Although excellent performance is obtained for large vocabulary tasks, tremendous memory consumption prohibits the use of LSTM LM in low-resource devices. The memory consumption mainly comes from the word embedding layer. In this paper, a novel binarized LSTM LM is proposed to address the problem. Words are encoded into binary vectors and other LSTM parameters are further binarized to achieve high memory compression. This is the first effort to investigate binary LSTM for large vocabulary LM. Experiments on both English and Chinese LM and ASR tasks showed that can achieve a compression ratio of 11.3 without any loss of LM and ASR performances and a compression ratio of 31.6 with acceptable minor performance degradation.

Conversational Memory Network for Emotion Recognition in Dyadic Dialogue Videos

Devamanyu Hazarika, Soujanya Poria, Amir Zadeh, Erik Cambria, Louis-Philippe Morency, and Roger Zimmermann

Emotion recognition in conversations is crucial for the development of empathetic machines. Present methods mostly ignore the role of inter-speaker dependency relations while classifying emotions in conversations. In this paper, we address recognizing utterance-level emotions in dyadic conversational videos. We propose a deep neural framework, termed Conversational Memory Network (CMN), which leverages contextual information from the conversation history. In particular, CMN uses multimodal approach comprising audio, visual and textual features with gated recurrent units to model past utterances of each speaker into memories. These memories are then merged using attention-based hops to capture inter-speaker dependencies. Experiments show a significant improvement of 3–4% in accuracy over the state of the art.

How Time Matters: Learning Time-Decay Attention for Contextual Spoken Language Understanding in Dialogues

Shang-Yu Su, Pei-Chieh Yuan, and Yun-Nung Chen

Spoken language understanding (SLU) is an essential component in conversational systems. Most SLU components treats each utterance independently, and then the following components aggregate the multi-turn information in the separate phases. In order to avoid error propagation and effectively utilize contexts, prior work leveraged history for contextual SLU. However, most previous models only paid attention to the related content in history utterances, ignoring their temporal information. In the dialogues, it is intuitive that the most recent utterances are more important than the least recent ones, in other words, time-aware attention should be in a decaying manner. Therefore, this paper designs and investigates various types of time-decay attention on the sentence-level and speaker-level, and further proposes a flexible universal time-decay attention mechanism. The experiments on the benchmark Dialogue State Tracking Challenge (DSTC4) dataset show that the proposed time-decay attention mechanisms significantly improve the state-of-the-art model for contextual understanding performance.

Role-specific Language Models for Processing Recorded Neuropsychological Exams

Tuka Al Hanai, Rhoda Au, and James Glass

Neuropsychological examinations are an important screening tool for the presence of cognitive conditions (e.g. Alzheimer's, Parkinson's Disease), and require a trained tester to conduct the exam through spoken interactions with the subject. While audio is relatively easy to record, it remains a challenge to automatically diarize (who spoke when?), decode (what did they say?), and assess a subject's cognitive health. This paper demonstrates a method to determine the cognitive health (impaired or not) of 92 subjects, from audio that was diarized using an automatic speech recognition system trained on TED talks and on the structured language used by testers and subjects. Using leave-one-out cross validation and logistic regression modeling we show that even with noisily decoded data (81% WER) we can still perform accurate enough diarization (0.02% confusion rate) to determine the cognitive state of a subject (0.76 AUC).

Slot-Gated Modeling for Joint Slot Filling and Intent Prediction

Chih-Wen Goo, Guang Gao, Yun-Kai Hsu, Chih-Li Huo, Tsung-Chieh Chen, Keng-Wei Hsu, and Yun-Nung Chen

Attention-based recurrent neural network models for joint intent detection and slot filling have achieved the state-of-the-art performance, while they have independent attention weights. Considering that slot and intent have the strong relationship, this paper proposes a slot gate that focuses on learning the relationship between intent and slot attention vectors in order to obtain better semantic frame results by the global optimization. The experiments show that our proposed model significantly improves sentence-level semantic frame accuracy with 4.2% and 1.9% relative improvement compared to the attentional model on benchmark ATIS and Snips datasets respectively.

Towards Understanding Text Factors in Oral Reading

Anastassia Loukina, Van Rynald T. Liceralde, and Beata Beigman Klebanov

Using a case study, we show that variation in oral reading rate across passages for professional narrators is consistent across readers and much of it can be explained using features of the texts being read. While text complexity is a poor predictor of the reading rate, a substantial share of variability can be explained by timing and story-based factors with performance reaching $r=0.75$ for unseen passages and narrator.

Posters: Vision, Robotics and Other Grounding 3

Time: 14:00–15:30

Location: Elite Hall B

Chair: Hannaneh Hajishirzi

An Evaluation of Image-Based Verb Prediction Models against Human Eye-Tracking Data

Spandana Gella and Frank Keller

Recent research in language and vision has developed models for predicting and disambiguating verbs from images. Here, we ask whether the predictions made by such models correspond to human intuitions about visual verbs. We show that the image regions a verb prediction model identifies as salient for a given verb correlate with the regions fixated by human observers performing a verb classification task.

Generating Bilingual Pragmatic Color References*Will Monroe, Jennifer Hu, Andrew Jong, and Christopher Potts*

Contextual influences on language often exhibit substantial cross-lingual regularities; for example, we are more verbose in situations that require finer distinctions. However, these regularities are sometimes obscured by semantic and syntactic differences. Using a newly-collected dataset of color reference games in Mandarin Chinese (which we release to the public), we confirm that a variety of constructions display the same sensitivity to contextual difficulty in Chinese and English. We then show that a neural speaker agent trained on bilingual data with a simple multitask learning approach displays more human-like patterns of context dependence and is more pragmatically informative than its monolingual Chinese counterpart. Moreover, this is not at the expense of language-specific semantic understanding: the resulting speaker model learns the different basic color term systems of English and Chinese (with noteworthy cross-lingual influences), and it can identify synonyms between the two languages using vector analogy operations on its output layer, despite having no exposure to parallel data.

Learning to Color from Language*Varun Manjunatha, Mohit Iyyer, Jordan Boyd-Graber, and Larry Davis*

Automatic colorization is the process of adding color to greyscale images. We condition this process on language, allowing end users to manipulate a colorized image by feeding in different captions. We present two different architectures for language-conditioned colorization, both of which produce more accurate and plausible colorizations than a language-agnostic version. Furthermore, we demonstrate through crowdsourced experiments that we can dramatically alter colorizations simply by manipulating descriptive color words in captions.

Learning with Latent Language*Jacob Andreas, Dan Klein, and Sergey Levine*

The named concepts and compositional operators present in natural language provide a rich source of information about the abstractions humans use to navigate the world. Can this linguistic background knowledge improve the generality and efficiency of learned classifiers and control policies? This paper aims to show that using the space of natural language strings as a parameter space is an effective way to capture natural task structure. In a pretraining phase, we learn a language interpretation model that transforms inputs (e.g. images) into outputs (e.g. labels) given natural language descriptions. To learn a new concept (e.g. a classifier), we search directly in the space of descriptions to minimize the interpreter's loss on training examples. Crucially, our models do not require language data to learn these concepts: language is used only in pretraining to impose structure on subsequent learning. Results on image classification, text editing, and reinforcement learning show that, in all settings, models with a linguistic parameterization outperform those without.

Object Counts! Bringing Explicit Detections Back into Image Captioning*Josiah Wang, Pranava Swaroop Madhyastha, and Lucia Specia*

The use of explicit object detectors as an intermediate step to image captioning – which used to constitute an essential stage in early work – is often bypassed in the currently dominant end-to-end approaches, where the language model is conditioned directly on a mid-level image embedding. We argue that explicit detections provide rich semantic information, and can thus be used as an interpretable representation to better understand why end-to-end image captioning systems work well. We provide an in-depth analysis of end-to-end image captioning by exploring a variety of cues that can be derived from such object detections. Our study reveals that end-to-end image captioning systems rely on matching image representations to generate captions, and that encoding the frequency, size and position of objects are complementary and all play a role in forming a good image representation. It also reveals that different object categories contribute in different ways towards image captioning.

Punny Captions: Witty Wordplay in Image Descriptions*Arjun Chandrasekaran, Devi Parikh, and Mohit Bansal*

Wit is a form of rich interaction that is often grounded in a specific situation (e.g., a comment in response to an event). In this work, we attempt to build computational models that can produce witty descriptions for a given image. Inspired by a cognitive account of humor appreciation, we employ linguistic wordplay, specifically puns, in image descriptions. We develop two approaches which involve retrieving witty descriptions for a given image from a large corpus of sentences, or generating them via an encoder-decoder neural network architecture. We compare our approach against meaningful baseline approaches via human studies and show substantial improvements. Moreover, in a Turing test style evaluation, people find the image descriptions generated by our model to be slightly wittier than human-written witty descriptions when the human is subject

to similar constraints as the model regarding word usage and style.

Quantifying the Visual Concreteness of Words and Topics in Multimodal Datasets

Jack Hessel, David Mimno, and Lillian Lee

Multimodal machine learning algorithms aim to learn visual-textual correspondences. Previous work suggests that concepts with concrete visual manifestations may be easier to learn than concepts with abstract ones. We give an algorithm for automatically computing the visual concreteness of words and topics within multimodal datasets. We apply the approach in four settings, ranging from image captions to images/text scraped from historical books. In addition to enabling explorations of concepts in multimodal datasets, our concreteness scores predict the capacity of machine learning algorithms to learn textual/visual relationships. We find that 1) concrete concepts are indeed easier to learn; 2) the large number of algorithms we consider have similar failure cases; 3) the precise positive relationship between concreteness and performance varies between datasets. We conclude with recommendations for using concreteness scores to facilitate future multimodal research.

Speaker Naming in Movies

Mahmoud Azab, Mingzhe Wang, Max Smith, Noriyuki Kojima, Jia Deng, and Rada Mihalcea

We propose a new model for speaker naming in movies that leverages visual, textual, and acoustic modalities in an unified optimization framework. To evaluate the performance of our model, we introduce a new dataset consisting of six episodes of the Big Bang Theory TV show and eighteen full movies covering different genres. Our experiments show that our multimodal model significantly outperforms several competitive baselines on the average weighted F-score metric. To demonstrate the effectiveness of our framework, we design an end-to-end memory network model that leverages our speaker naming model and achieves state-of-the-art results on the subtitles task of the MovieQA 2017 Challenge.

Stacking with Auxiliary Features for Visual Question Answering

Nazneen Fatema Rajani and Raymond Mooney

Visual Question Answering (VQA) is a well-known and challenging task that requires systems to jointly reason about natural language and vision. Deep learning models in various forms have been the standard for solving VQA. However, some of these VQA models are better at certain types of image-question pairs than other models. Ensembling VQA models intelligently to leverage their diverse expertise is, therefore, advantageous. Stacking With Auxiliary Features (SWAF) is an intelligent ensembling technique which learns to combine the results of multiple models using features of the current problem as context. We propose four categories of auxiliary features for ensembling for VQA. Three out of the four categories of features can be inferred from an image-question pair and do not require querying the component models. The fourth category of auxiliary features uses model-specific explanations. In this paper, we describe how we use these various categories of auxiliary features to improve performance for VQA. Using SWAF to effectively ensemble three recent systems, we obtain a new state-of-the-art. Our work also highlights the advantages of explainable AI models.

The Emergence of Semantics in Neural Network Representations of Visual Information

Dhanush Dharmaretnam and Alona Fyshe

Word vector models learn about semantics through corpora. Convolutional Neural Networks (CNNs) can learn about semantics through images. At the most abstract level, some of the information in these models must be shared, as they model the same real-world phenomena. Here we employ techniques previously used to detect semantic representations in the human brain to detect semantic representations in CNNs. We show the accumulation of semantic information in the layers of the CNN, and discover that, for misclassified images, the correct class can be recovered in intermediate layers of a CNN.

Visual Referring Expression Recognition: What Do Systems Actually Learn?

Volkan Cirik, Louis-Philippe Morency, and Taylor Berg-Kirkpatrick

We present an empirical analysis of state-of-the-art systems for referring expression recognition – the task of identifying the object in an image referred to by a natural language expression – with the goal of gaining insight into how these systems reason about language and vision. Surprisingly, we find strong evidence that even sophisticated and linguistically-motivated models for this task may ignore linguistic structure, instead relying on shallow correlations introduced by unintended biases in the data selection and annotation process. For example, we show that a system trained and tested on the input image without the input referring expression can achieve a precision of 71.2% in top-2 predictions. Furthermore, a system that predicts only the object category given the input can achieve a precision of 84.2% in top-2 predictions. These surprisingly positive results for what should be deficient prediction scenarios suggest that careful analysis of what our models are learning – and further, how our data is constructed – is critical as we seek to make substantive progress on

grounded language tasks.

Visually Guided Spatial Relation Extraction from Text

Taher Rahgooy, Umar Manzoor, and Parisa Kordjamshidi

Extraction of spatial relations from sentences with complex/nesting relationships is very challenging as often needs resolving inherent semantic ambiguities. We seek help from visual modality to fill the information gap in the text modality and resolve spatial semantic ambiguities. We use various recent vision and language datasets and techniques to train inter-modality alignment models, visual relationship classifiers and propose a novel global inference model to integrate these components into our structured output prediction model for spatial role and relation extraction. Our global inference model enables us to utilize the visual and geometric relationships between objects and improves the state-of-art results of spatial information extraction from text.

Watch, Listen, and Describe: Globally and Locally Aligned Cross-Modal Attentions for Video Captioning

Xin Wang, Yuan-Fang Wang, and William Yang Wang

A major challenge for video captioning is to combine audio and visual cues. Existing multi-modal fusion methods have shown encouraging results in video understanding. However, the temporal structures of multiple modalities at different granularities are rarely explored, and how to selectively fuse the multi-modal representations at different levels of details remains uncharted. In this paper, we propose a novel hierarchically aligned cross-modal attention (HACA) framework to learn and selectively fuse both global and local temporal dynamics of different modalities. Furthermore, for the first time, we validate the superior performance of the deep audio features on the video captioning task. Finally, our HACA model significantly outperforms the previous best systems and achieves new state-of-the-art results on the widely used MSR-VTT dataset.

Generating Image Captions in Arabic using Root-Word Based Recurrent Neural Networks and Deep Neural Networks (SRW)

Vasu Jindal

Image caption generation has gathered widespread interest in the artificial intelligence community. Automatic generation of an image description requires both computer vision and natural language processing techniques. While, there has been advanced research in the English caption generation, research on generating Arabic descriptions of an image is extremely limited. Semitic languages like Arabic are heavily influenced by root-words. We leverage this critical dependency of Arabic to generate captions of an image directly in Arabic using root-word based Recurrent Neural Network and Deep Neural Networks. Experimental results on dataset from various Middle Eastern newspaper websites allow us to report the first BLEU score for direct Arabic caption generation. We also compare the results of our approach with BLEU score captions generated in English and translated in Arabic. Experimental results confirm that generating image captions using root-words directly in Arabic significantly outperforms the English-Arabic translated captions using state-of-the-art methods.

Demos

Time: 14:00–15:30

Location: Elite Hall B

SMILEE: Symmetric Multi-modal Interactions with Language-gesture Enabled (AI) Embodiment

Sujeong Kim, David Salter, Luke DeLucia, Kilho Son, Mohamed R. Amer, and Amir Tamrakar

We demonstrate an intelligent conversational agent system designed for advancing human-machine collaborative tasks. The agent is able to interpret a user's communicative intent from both their verbal utterances and non-verbal behaviors, such as gestures. The agent is also itself able to communicate both with natural language and gestures, through its embodiment as an avatar thus facilitating natural symmetric multi-modal interactions. We demonstrate two intelligent agents with specialized skills in the Blocks World as use-cases of our system.

Decision Conversations Decoded

Lea Deleris, Debasis Ganguly, Killian Levacher, Martin Stephenson, and Francesca Bonin

We describe the vision and current version of a Natural Language Processing system aimed at group decision making facilitation. Borrowing from the scientific field of Decision Analysis, its essential role is to identify alternatives and criteria associated with a given decision, to keep track of who proposed them and of the expressed sentiment towards them. Based on this information, the system can help identify agreement and dissent or recommend an alternative. Overall, it seeks to help a group reach a decision in a natural yet auditable

fashion.

Sounding Board: A User-Centric and Content-Driven Social Chatbot

Hao Fang, Hao Cheng, Maarten Sap, Elizabeth Clark, Ari Holtzman, Yejin Choi, Noah A. Smith, and Mari Ostendorf

We present Sounding Board, a social chatbot that won the 2017 Amazon Alexa Prize. The system architecture consists of several components including spoken language processing, dialogue management, language generation, and content management, with emphasis on user-centric and content-driven design. We also share insights gained from large-scale online logs based on 160,000 conversations with real-world users.

Session 12 Overview – Monday, June 4, 2018

Outstanding Paper Session (sponsored by Amazon)

Empire Ballroom

Deep Contextualized Word Representations

M. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer

Learning to Map Context-Dependent Sentences to Executable Formal Queries

A. Suhr, S. Iyer, and Y. Artzi

Neural Text Generation in Stories Using Entity Representations as Context

E. Clark, Y. Ji, and N. A. Smith

Recurrent Neural Networks as Weighted Language Recognizers

Y. Chen, S. Gilroy, A. Maletti, J. May, and K. Knight

17:54 17:36 17:18 17:00

Session 12

Outstanding Paper Session (sponsored by Amazon)

Empire A

Chair: Joel Tetreault

Deep Contextualized Word Representations

Matthew Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer 17:00–17:17

We introduce a new type of deep contextualized word representation that models both (1) complex characteristics of word use (e.g., syntax and semantics), and (2) how these uses vary across linguistic contexts (i.e., to model polysemy). Our word vectors are learned functions of the internal states of a deep bidirectional language model (biLM), which is pre-trained on a large text corpus. We show that these representations can be easily added to existing models and significantly improve the state of the art across six challenging NLP problems, including question answering, textual entailment and sentiment analysis. We also present an analysis showing that exposing the deep internals of the pre-trained network is crucial, allowing downstream models to mix different types of semi-supervision signals.

Learning to Map Context-Dependent Sentences to Executable Formal Queries

Alane Suhr, Srinivasan Iyer, and Yoav Artzi 17:18–17:35

We propose a context-dependent model to map utterances within an interaction to executable formal queries. To incorporate interaction history, the model maintains an interaction-level encoder that updates after each turn, and can copy sub-sequences of previously predicted queries during generation. Our approach combines implicit and explicit modeling of references between utterances. We evaluate our model on the ATIS flight planning interactions, and demonstrate the benefits of modeling context and explicit references.

Neural Text Generation in Stories Using Entity Representations as Context

Elizabeth Clark, Yangfeng Ji, and Noah A. Smith 17:36–17:53

We introduce an approach to neural text generation that explicitly represents entities mentioned in the text. Entity representations are vectors that are updated as the text proceeds; they are designed specifically for narrative text like fiction or news stories. Our experiments demonstrate that modeling entities offers a benefit in two automatic evaluations: mention generation (in which a model chooses which entity to mention next and which words to use in the mention) and selection between a correct next sentence and a distractor from later in the same story. We also conduct a human evaluation on automatically generated text in story contexts; this study supports our emphasis on entities and suggests directions for further research.

Recurrent Neural Networks as Weighted Language Recognizers

Yining Chen, Sorcha Gilroy, Andreas Maletti, Jonathan May, and Kevin Knight 17:54–18:12

We investigate the computational complexity of various problems for simple recurrent neural networks (RNNs) as formal models for recognizing weighted languages. We focus on the single-layer, ReLU-activation, rational-weight RNNs with softmax, which are commonly used in natural language processing applications. We show that most problems for such RNNs are undecidable, including consistency, equivalence, minimization, and the determination of the highest-weighted string. However, for consistent RNNs the last problem becomes decidable, although the solution length can surpass all computable bounds. If additionally the string is limited to polynomial length, the problem becomes NP-complete. In summary, this shows that approximations and heuristic algorithms are necessary in practical applications of those RNNs.

6

Workshops and Collocated Events

Tuesday–Wednesday

Strand 10	SemEval: 12th International Workshop on Semantic Evaluation	p.152
Strand 12B	*SEM: Seventh Joint Conference On Lexical And Computational Semantics	p.165

Tuesday

Strand 11	GenDeep: Generalization in the Age of Deep Learning	p.168
Strand 12A	BEA: 13th Workshop on Innovative Use of NLP for Building Educational Applications	p.169
Bolden 1	Second ACL Workshop on Ethics in Natural Language Processing	p.173
Bolden 5	CLPsych: Fifth Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality	p.174
Bolden 4	SemBEaR: Computational Semantics Beyond Events and Roles	p.176
Bolden 3	Style-Var: 2nd Workshop on Stylistic Variation	p.177
Bolden 6	Story-NLP: Workshop on Storytelling	p.178

Wednesday

Bolden 6	SCLeM: Subword and Character LEvel Models in NLP	p.179
Bolden 1	TextGraphs: Twelfth Workshop on Graph-Based Methods for Natural Language Processing	p.180
Bolden 2	FigLang: Workshop on Figurative Language Processing	p.181
Strand 11	PEOPLES: Second Workshop on Computational Modeling of PEople's Opinions, PersonaLity and Emotions in Social Media	p.183
Bolden 3	CRAC: Workshop on Computational models of Reference, Anaphora and Coreference	p.185
Bolden 4	SpLU: First International Workshop on Spatial Language Understanding	p.186

SemEval: 12th International Workshop on Semantic Evaluation

Organizers: *Marianna Apidianaki, Saif Mohammad, Jonathan May, Ekaterina Shutova, Steven Bethard, and Marine Carpuat*

Venue: Strand 10

Tuesday, June 5, 2018

08:00–17:00 **NAACL Registration**

09:00–09:30 **Welcome / Opening Remarks**

09:30–10:30 **Invited Talk: Why Should We Care about Linguistics? (Ellie Pavlick)**

10:30–11:00 **Coffee**

11:00–12:30 **Tasks 1, 2 and 3**

11:00–11:15 SemEval-2018 Task 1: Affect in Tweets
Saif Mohammad, Felipe Bravo-Marquez, Mohammad Salameh, and Svetlana Kiritchenko

11:15–11:30 SeerNet at SemEval-2018 Task 1: Domain Adaptation for Affect in Tweets
Venkatesh Duppada, Royal Jain, and Sushant Hiray

11:30–11:45 SemEval 2018 Task 2: Multilingual Emoji Prediction
Francesco Barbieri, Jose Camacho-Collados, Francesco Ronzano, Luis Espinosa Anke, Miguel Ballesteros, Valerio Basile, Viviana Patti, and Horacio Saggion

11:45–12:00 Tübingen-Oslo at SemEval-2018 Task 2: SVMs perform better than RNNs in Emoji Prediction
Çağrı Çöltekin and Taraka Rama

12:00–12:15 SemEval-2018 Task 3: Irony Detection in English Tweets
Cynthia Van Hee, Els Lefever, and Veronique Hoste

12:15–12:30 THU_NGN at SemEval-2018 Task 3: Tweet Irony Detection with Densely connected LSTM and Multi-task Learning
Chuhan Wu, Fangzhao Wu, Sixing Wu, Junxin Liu, Zhigang Yuan, and Yongfeng Huang

12:30–14:00 **Lunch**

14:00–15:30 **Tasks 4, 5 and 6**

14:00–14:15 SemEval 2018 Task 4: Character Identification on Multiparty Dialogues
Jinho D. Choi and Henry Y. Chen

14:15–14:30 AMORE-UPF at SemEval-2018 Task 4: BiLSTM with Entity Library
Laura Aina, Carina Silberer, Ionut-Teodor Sorodoc, Matthijs Westera, and Gemma Boleda

14:30–14:45 SemEval-2018 Task 5: Counting Events and Participants in the Long Tail
Marten Postma, Filip Ilievski, and Piek Vossen

14:45–15:00 KOI at SemEval-2018 Task 5: Building Knowledge Graph of Incidents
Paramita Mirza, Fariz Darari, and Rahmad Mahendra

15:00–15:15 SemEval 2018 Task 6: Parsing Time Normalizations
Egoitz Laparra, Dongfang Xu, Ahmed Elsayed, Steven Bethard, and Martha Palmer

- 15:15–15:30 Chrono at SemEval-2018 Task 6: A System for Normalizing Temporal Expressions
Amy Olex, Luke Maffey, Nicholas Morgan, and Bridget McInnes
- 15:30–16:00 **Coffee**
- 16:00–16:30 **Discussion**
- 16:30–17:30 **Poster Session**
- 16:30–17:30 NEUROSENT-PDI at SemEval-2018 Task 1: Leveraging a Multi-Domain Sentiment Model for Inferring Polarity in Micro-blog Text
Mauro Dragomi
- 16:30–17:30 FOI DSS at SemEval-2018 Task 1: Combining LSTM States, Embeddings, and Lexical Features for Affect Analysis
Maja Karasalo, Mattias Nilsson, Magnus Rosell, and Ulrika Wickenberg Bolin
- 16:30–17:30 NLPZZX at SemEval-2018 Task 1: Using Ensemble Method for Emotion and Sentiment Intensity Determination
Zhengxin Zhang, Qimin Zhou, and Hao Wu
- 16:30–17:30 LT3 at SemEval-2018 Task 1: A classifier chain to detect emotions in tweets
Luna De Bruyne, Orphee De Clercq, and Veronique Hoste
- 16:30–17:30 SINAI at SemEval-2018 Task 1: Emotion Recognition in Tweets
Flor Miriam Plaza del Arco, Salud María Jiménez-Zafra, Maite Martin, and L. Alfonso Urena Lopez
- 16:30–17:30 UWB at SemEval-2018 Task 1: Emotion Intensity Detection in Tweets
Pavel Přibáň, Tomáš Hercig, and Ladislav Lenc
- 16:30–17:30 AttnConvnet at SemEval-2018 Task 1: Attention-based Convolutional Neural Networks for Multi-label Emotion Classification
Yanghoon Kim, Hwanhee Lee, and Kyomin Jung
- 16:30–17:30 INGEOTEC at SemEval-2018 Task 1: EvoMSA and μTC for Sentiment Analysis
Mario Graff, Sabino Miranda-Jiménez, Eric S. Tellez, and Daniela Moctezuma
- 16:30–17:30 Epita at SemEval-2018 Task 1: Sentiment Analysis Using Transfer Learning Approach
Guillaume Daval-Frerot, Abdesselam Boucheikif, and Anatole Moreau
- 16:30–17:30 KDE-AFFECT at SemEval-2018 Task 1: Estimation of Affects in Tweet by Using Convolutional Neural Network for n-gram
Masaki Aono and Shinnosuke Himeno
- 16:30–17:30 RNN for Affects at SemEval-2018 Task 1: Formulating Affect Identification as a Binary Classification Problem
Aysu Ezen-Can and Ethem F. Can
- 16:30–17:30 Tw-StAR at SemEval-2018 Task 1: Preprocessing Impact on Multi-label Emotion Classification
Hala Mulki, Chedi Bechikh Ali, Hatem Haddad, and Ismail Babaoglu
- 16:30–17:30 DL Team at SemEval-2018 Task 1: Tweet Affect Detection using Sentiment Lexicons and Embeddings
Dmitry Kravchenko and Lidia Pivovarova
- 16:30–17:30 EmoIntens Tracker at SemEval-2018 Task 1: Emotional Intensity Levels in #Tweets
Ramona-Andreea Turcu, Sandra Maria Amarandei, Iuliana-Alexandra Fleşcan-Lovin-Arseni, Daniela Gifu, and Diana Trandabat

- 16:30–17:30 uOttawa at SemEval-2018 Task 1: Self-Attentive Hybrid GRU-Based Network
Ahmed Husseini Orabi, Mahmoud Husseini Orabi, Diana Inkpen, and David Van Buwaene
- 16:30–17:30 THU_NGN at SemEval-2018 Task 1: Fine-grained Tweet Sentiment Intensity Analysis with Attention CNN-LSTM
Chuhan Wu, Fangzhao Wu, Junxin Liu, Zhigang Yuan, Sixing Wu, and Yongfeng Huang
- 16:30–17:30 EiTAKA at SemEval-2018 Task 1: An Ensemble of N-Channels ConvNet and XGboost Regressors for Emotion Analysis of Tweets
Mohammed Jabreel and Antonio Moreno
- 16:30–17:30 CENTEMENT at SemEval-2018 Task 1: Classification of Tweets using Multiple Thresholds with Self-correction and Weighted Conditional Probabilities
Tariq Ahmad, Allan Ramsay, and Hanady Ahmed
- 16:30–17:30 Yuan at SemEval-2018 Task 1: Tweets Emotion Intensity Prediction using Ensemble Recurrent Neural Network
Min Wang and Xiaobing Zhou
- 16:30–17:30 AffecThor at SemEval-2018 Task 1: A cross-linguistic approach to sentiment intensity quantification in tweets
Mostafa Abdou, Artur Kulmizev, and Joan Ginés i Ametller
- 16:30–17:30 Amobee at SemEval-2018 Task 1: GRU Neural Network with a CNN Attention Mechanism for Sentiment Classification
Alon Rozental and Daniel Fleischer
- 16:30–17:30 deepSA2018 at SemEval-2018 Task 1: Multi-task Learning of Different Label for Affect in Tweets
Zi Yuan Gao and Chia-Ping Chen
- 16:30–17:30 ECNU at SemEval-2018 Task 1: Emotion Intensity Prediction Using Effective Features and Machine Learning Models
Huimin Xu, Man Lan, and Yuanbin Wu
- 16:30–17:30 EMA at SemEval-2018 Task 1: Emotion Mining for Arabic
Gilbert Badaro, Obeida El Jundi, Alaa Khaddaj, Alaa Maarouf, Raslan Kain, Hazem Hajj, and Wassim El-Hajj
- 16:30–17:30 NTUA-SLP at SemEval-2018 Task 1: Predicting Affective Content in Tweets with Deep Attentive RNNs and Transfer Learning
Christos Baziots, Athanasiou Nikolaos, Alexandra Chronopoulou, Athanasia Kolovou, Georgios Paraskevopoulos, Nikolaos Ellinas, Shrikanth Narayanan, and Alexandros Potamianos
- 16:30–17:30 CrystalFeel at SemEval-2018 Task 1: Understanding and Detecting Emotion Intensity using Affective Lexicons
Raj Kumar Gupta and Yingping Yang
- 16:30–17:30 PlusEmo2Vec at SemEval-2018 Task 1: Exploiting emotion knowledge from emoji and #hashtags
Ji Ho Park, Peng Xu, and Pascale Fung
- 16:30–17:30 YNU-HPCC at SemEval-2018 Task 1: BiLSTM with Attention based Sentiment Analysis for Affect in Tweets
You Zhang, Jin Wang, and Xuejie Zhang
- 16:30–17:30 UG18 at SemEval-2018 Task 1: Generating Additional Training Data for Predicting Emotion Intensity in Spanish
Marloes Kuijper, Mike van Lenthe, and Rik van Noord
- 16:30–17:30 ISCLAB at SemEval-2018 Task 1: UIR-Miner for Affect in Tweets
Meng Li, Zhenyuan Dong, Zhihao Fan, Kongming Meng, Jinghua Cao, Guanqi Ding, Yuhan Liu, Jiawei Shan, and Binyang Li

- 16:30–17:30 TCS Research at SemEval-2018 Task 1: Learning Robust Representations using Multi-Attention Architecture
Hardik Meisheri and Lipika Dey
- 16:30–17:30 DMCB at SemEval-2018 Task 1: Transfer Learning of Sentiment Classification Using Group LSTM for Emotion Intensity prediction
Youngmin Kim and Hyunju Lee
- 16:30–17:30 DeepMiner at SemEval-2018 Task 1: Emotion Intensity Recognition Using Deep Representation Learning
Habibeh Naderi, Behrouz Haji Soleimani, Saif Mohammad, Svetlana Kiritchenko, and Stan Matwin
- 16:30–17:30 Zewen at SemEval-2018 Task 1: An Ensemble Model for Affect Prediction in Tweets
Zewen Chi, Heyan Huang, Jiangui Chen, Hao Wu, and Ran Wei
- 16:30–17:30 Amrita_student at SemEval-2018 Task 1: Distributed Representation of Social Media Text for Affects in Tweets
Nidhin A Unnithan, Shalini K, Barathi Ganesh H. B., Anand Kumar M, and Soman K P
- 16:30–17:30 SSN MLRG1 at SemEval-2018 Task 1: Emotion and Sentiment Intensity Detection Using Rule Based Feature Selection
Angel Deborah S, Rajalakshmi S, S Milton Rajendram, and Mirnalinee T T
- 16:30–17:30 CENNLP at SemEval-2018 Task 1: Constrained Vector Space Model in Affects in Tweets
Naveen J R, Barathi Ganesh H. B., Anand Kumar M, and Soman K P
- 16:30–17:30 TeamCEN at SemEval-2018 Task 1: Global Vectors Representation in Emotion Detection
Anon George, Barathi Ganesh H. B., Anand Kumar M, and Soman K P
- 16:30–17:30 IIT Delhi at SemEval-2018 Task 1 : Emotion Intensity Prediction
Bhaskar Kotakonda, Prashanth Gowda, and Brejesh Lall
- 16:30–17:30 Mutux at SemEval-2018 Task 1: Exploring Impacts of Context Information On Emotion Detection
Pan Du and Jian-Yun Nie
- 16:30–17:30 TeamUNCC at SemEval-2018 Task 1: Emotion Detection in English and Arabic Tweets using Deep Learning
Malak Abdullah and Samira Shaikh
- 16:30–17:30 RIDDL at SemEval-2018 Task 1: Rage Intensity Detection with Deep Learning
Venkatesh Elango and Karan Uppal
- 16:30–17:30 ARB-SEN at SemEval-2018 Task1: A New Set of Features for Enhancing the Sentiment Intensity Prediction in Arabic Tweets
El Moatez Billah Nagoudi
- 16:30–17:30 psyML at SemEval-2018 Task 1: Transfer Learning for Sentiment and Emotion Analysis
Grace Gee and Eugene Wang
- 16:30–17:30 UIUC at SemEval-2018 Task 1: Recognizing Affect with Ensemble Models
Abhishek Avinash Narwekar and Roxana Girju
- 16:30–17:30 KU-MTL at SemEval-2018 Task 1: Multi-task Identification of Affect in Tweets
Thomas Nyegaard-Signori, Casper Veistrup Helms, Johannes Bjerva, and Isabelle Augenstein
- 16:30–17:30 EmoNLP at SemEval-2018 Task 2: English Emoji Prediction with Gradient Boosting Regression Tree Method and Bidirectional LSTM
Man Liu

- 16:30–17:30 UMDSub at SemEval-2018 Task 2: Multilingual Emoji Prediction
Multi-channel Convolutional Neural Network on Subword Embedding
Zhenduo Wang and Ted Pedersen
- 16:30–17:30 UMDuluth-CS8761 at SemEval-2018 Task 2: Emojis: Too many Choices?
Jonathan Beaulieu and Dennis Asamoah Owusu
- 16:30–17:30 The Dabblers at SemEval-2018 Task 2: Multilingual Emoji Prediction
Larisa Alexa, Alina Lorent, Daniela Gifu, and Diana Trandabat
- 16:30–17:30 THU_NGN at SemEval-2018 Task 2: Residual CNN-LSTM Network with Attention for English Emoji Prediction
Chuhan Wu, Fangzhao Wu, Sixing Wu, Zhigang Yuan, Junxin Liu, and Yongfeng Huang
- 16:30–17:30 #TeamINF at SemEval-2018 Task 2: Emoji Prediction in Tweets
Alison Ribeiro and Nádia Silva
- 16:30–17:30 EICA Team at SemEval-2018 Task 2: Semantic and Metadata-based Features for Multilingual Emoji Prediction
Yufei Xie and Qingqing Song
- 16:30–17:30 EmojIIt at SemEval-2018 Task 2: An Effective Attention-Based Recurrent Neural Network Model for Emoji Prediction with Characters Gated Words
Chen Shiyun, Wang Maoquan, and He Liang
- 16:30–17:30 Peperomia at SemEval-2018 Task 2: Vector Similarity Based Approach for Emoji Prediction
Jing Chen, Dechuan Yang, Xilian Li, Wei Chen, and Tengjiao Wang
- 16:30–17:30 ECNU at SemEval-2018 Task 2: Leverage Traditional NLP Features and Neural Networks Methods to Address Twitter Emoji Prediction Task
Xingwu Lu, Xin Mao, Man Lan, and Yuanbin Wu
- 16:30–17:30 NTUA-SLP at SemEval-2018 Task 2: Predicting Emojis using RNNs with Context-aware Attention
Christos Baziotis, Athanasiou Nikolaos, Athanasia Kolovou, Georgios Paraskevopoulos, Nikolaos Ellinas, and Alexandros Potamianos
- 16:30–17:30 Hatching Chick at SemEval-2018 Task 2: Multilingual Emoji Prediction
Joël Coster, Reinder Gerard van Dalen, and Nathalie Adriënne Jacqueline Stierman
- 16:30–17:30 EPUTION at SemEval-2018 Task 2: Emoji Prediction with User Adaption
Liyuan Zhou, Qiongkai Xu, Hanna Suominen, and Tom Gedeon
- 16:30–17:30 PickleTeam! at SemEval-2018 Task 2: English and Spanish Emoji Prediction from Tweets
Daphne Groot, Rémon Kruizinga, Hennie Veldthuis, Simon de Wit, and Hessel Haagsma
- 16:30–17:30 YNU-HPCC at SemEval-2018 Task 2: Multi-ensemble Bi-GRU Model with Attention Mechanism for Multilingual Emoji Prediction
Nan Wang, Jin Wang, and Xuejie Zhang
- 16:30–17:30 DUTH at SemEval-2018 Task 2: Emoji Prediction in Tweets
Dimitrios Effrosynidis, Georgios Peikos, Symeon Symeonidis, and Avi Arampatzis
- 16:30–17:30 TAJJEB at SemEval-2018 Task 2: Traditional Approaches Just Do the Job with Emoji Prediction
Angelo Basile and Kenny W. Lino
- 16:30–17:30 SyntNN at SemEval-2018 Task 2: is Syntax Useful for Emoji Prediction? Embedding Syntactic Trees in Multi Layer Perceptrons
Fabio Massimo Zanzotto and Andrea Santilli
- 16:30–17:30 Duluth UROP at SemEval-2018 Task 2: Multilingual Emoji Prediction with Ensemble Learning and Oversampling
Shuning Jin and Ted Pedersen

- 16:30–17:30 CENNLP at SemEval-2018 Task 2: Enhanced Distributed Representation of Text using Target Classes for Emoji Prediction Representation
Naveen J R, Hariharan V, Barathi Ganesh H. B., Anand Kumar M, and Soman K P
- 16:30–17:30 Manchester Metropolitan at SemEval-2018 Task 2: Random Forest with an Ensemble of Features for Predicting Emoji in Tweets
Luciano Gerber and Matthew Shardlow
- 16:30–17:30 Tweety at SemEval-2018 Task 2: Predicting Emojis using Hierarchical Attention Neural Networks and Support Vector Machine
Daniel Kopev, Atanas Atanasov, Dimitrina Zlatkova, Momchil Hardalov, Ivan Koychev, Ivelina Nikolova, and Galia Angelova
- 16:30–17:30 LIS at SemEval-2018 Task 2: Mixing Word Embeddings and Bag of Features for Multilingual Emoji Prediction
Gaël Guibon, Magalie Ochs, and Patrice Bellot
- 16:30–17:30 ALANIS at SemEval-2018 Task 3: A Feature Engineering Approach to Irony Detection in English Tweets
Kevin Swanberg, Madiha Mirza, Ted Pedersen, and Zhenduo Wang
- 16:30–17:30 NEUROSENT-PDI at SemEval-2018 Task 3: Understanding Irony in Social Networks Through a Multi-Domain Sentiment Model
Mauro Dragomi
- 16:30–17:30 UWB at SemEval-2018 Task 3: Irony detection in English tweets
Tomáš Hercig
- 16:30–17:30 NIHARIO at SemEval-2018 Task 3: A Simple and Accurate Neural Network Model for Irony Detection in Twitter
Thanh Vu, Dat Quoc Nguyen, Xuan-Son Vu, Dai Quoc Nguyen, Michael Catt, and Michael Trenell
- 16:30–17:30 LDR at SemEval-2018 Task 3: A Low Dimensional Text Representation for Irony Detection
Bilal Ghanem, Francisco Rangel, and Paolo Rosso
- 16:30–17:30 IIDYT at SemEval-2018 Task 3: Irony detection in English tweets
Edison Marrese-Taylor, Suzana Ilic, Jorge Balazs, Helmut Prendinger, and Yutaka Matsuo
- 16:30–17:30 PunFields at SemEval-2018 Task 3: Detecting Irony by Tools of Humor Analysis
Elena Mikhalkova, Yuri Karyakin, Alexander Voronov, Dmitry Grigoriev, and Artem Leoznov
- 16:30–17:30 HashCount at SemEval-2018 Task 3: Concatenative Featurization of Tweet and Hashtags for Irony Detection
Won Ik Cho, Woo Hyun Kang, and Nam Soo Kim
- 16:30–17:30 WLV at SemEval-2018 Task 3: Dissecting Tweets in Search of Irony
Omnid Rohanian, Shiva Taslimipoor, Richard Evans, and Ruslan Mitkov
- 16:30–17:30 Random Decision Syntax Trees at SemEval-2018 Task 3: LSTMs and Sentiment Scores for Irony Detection
Aidan San
- 16:30–17:30 ELiRF-UPV at SemEval-2018 Tasks 1 and 3: Affect and Irony Detection in Tweets
José-Angel González, Lluís-F. Hurtado, and Ferran Pla
- 16:30–17:30 IronyMagnet at SemEval-2018 Task 3: A Siamese network for Irony detection in Social media
Aniruddha Ghosh and Tony Veale
- 16:30–17:30 CTSys at SemEval-2018 Task 3: Irony in Tweets
Myan Sherif, Sherine Mamdouh, and Wegdan Ghazi

- 16:30–17:30 Irony Detector at SemEval-2018 Task 3: Irony Detection in English Tweets using Word Graph
Usman Ahmed, Lubna Zafar, Faiza Qayyum, and Muhammad Arshad Islam
- 16:30–17:30 Lancaster at SemEval-2018 Task 3: Investigating Ironic Features in English Tweets
Edward Dearden and Alistair Baron
- 16:30–17:30 INAOE-UPV at SemEval-2018 Task 3: An Ensemble Approach for Irony Detection in Twitter
Delia Irazú Hernández Farías, Fernando Sánchez-Vega, Manuel Montes, and Paolo Rosso
- 16:30–17:30 ECNU at SemEval-2018 Task 3: Exploration on Irony Detection from Tweets via Machine Learning and Deep Learning Methods
Zhenghang Yin, Feixiang Wang, Man Lan, and Wenting Wang
- 16:30–17:30 KLUEncorn at SemEval-2018 Task 3: A Naive Approach to Irony Detection
Luise Dürrlich
- 16:30–17:30 NTUA-SLP at SemEval-2018 Task 3: Tracking Ironic Tweets using Ensembles of Word and Character Level Attentive RNNs
Christos Baziotis, Athanasiou Nikolaos, Pinelopi Papalampidi, Athanasia Kolovou, Georgios Paraskevopoulos, Nikolaos Ellinas, and Alexandros Potamianos
- 16:30–17:30 YNU-HPCC at SemEval-2018 Task 3: Ensemble Neural Network Models for Irony Detection on Twitter
Bo Peng, Jin Wang, and Xuejie Zhang
- 16:30–17:30 Binarizer at SemEval-2018 Task 3: Parsing dependency and deep learning for irony detection
Nishant Nikhil and Muktabh Mayank Srivastava
- 16:30–17:30 SSN MLRG1 at SemEval-2018 Task 3: Irony Detection in English Tweets Using MultiLayer Perceptron
Rajalakshmi S, Angel Deborah S, S Milton Rajendram, and Mirnalinee T T
- 16:30–17:30 NLPRL-IITBHU at SemEval-2018 Task 3: Combining Linguistic Features and Emoji pre-trained CNN for Irony Detection in Tweets
Harsh Rangwani, Devang Kulshreshtha, and Anil Kumar Singh
- 16:30–17:30 ValenTO at SemEval-2018 Task 3: Exploring the Role of Affective Content for Detecting Irony in English Tweets
Delia Irazú Hernández Farías, Viviana Patti, and Paolo Rosso
- 16:30–17:30 #NonDicevoSulSerio at SemEval-2018 Task 3: Exploiting Emojis and Affective Content for Irony Detection in English Tweets
Endang Wahyu Pamungkas and Viviana Patti
- 16:30–17:30 KNU CI System at SemEval-2018 Task4: Character Identification by Solving Sequence-Labeling Problem
Cheoneum Park, Heejun Song, and Changki Lee
- 16:30–17:30 NewsReader at SemEval-2018 Task 5: Counting events by reasoning over event-centric-knowledge-graphs
Piek Vossen
- 16:30–17:30 FEUP at SemEval-2018 Task 5: An Experimental Study of a Question Answering System
Carla Abreu and Eugénio Oliveira
- 16:30–17:30 NAI-SEA at SemEval-2018 Task 5: An Event Search System
Yingchi Liu, Quanzhi Li, and Luo Si

Wednesday, June 6, 2018

08:00–16:00 NAACL Registration

09:00–09:30 SemEval 2019 Tasks

09:30–10:30 State of SemEval Discussion

10:30–11:00 Coffee

11:00–12:30 Tasks 7, 8 and 9

11:00–11:15 SemEval-2018 Task 7: Semantic Relation Extraction and Classification in Scientific Papers

*Kata Gábor, Davide Buscaldi, Anne-Kathrin Schumann,
Behrang QasemiZadeh, Haifa Zargayouna, and Thierry Charnois*

11:15–11:30 ETH-DS3Lab at SemEval-2018 Task 7: Effectively Combining Recurrent and Convolutional Neural Networks for Relation Classification and Extraction

Jonathan Rotsztejn, Nora Hollenstein, and Ce Zhang

11:30–11:45 SemEval-2018 Task 8: Semantic Extraction from CybersecUrity REports using Natural Language Processing (SecureNLP)

Peter Phandi, Amila Silva, and Wei Lu

11:45–12:00 DM_NLP at SemEval-2018 Task 8: neural sequence labeling with linguistic features

Chunping Ma, Huafei Zheng, Pengjun Xie, Chen Li, Linlin Li, and Luo Si

12:00–12:15 SemEval-2018 Task 9: Hypernym Discovery

*Jose Camacho-Collados, Claudio Delli Bovi, Luis Espinosa Anke,
Sergio Oramas, Tommaso Pasini, Enrico Santus, Vered Shwartz,
Roberto Navigli, and Horacio Saggion*

12:15–12:30 CRIM at SemEval-2018 Task 9: A Hybrid Approach to Hypernym Discovery

Gabriel Bernier-Colborne and Caroline Barriere

12:30–14:00 Lunch

14:00–15:30 Tasks 10, 11 and 12

14:00–14:15 SemEval-2018 Task 10: Capturing Discriminative Attributes

Alicia Krebs, Alessandro Lenci, and Denis Paperno

14:15–14:30 SUNNYNLP at SemEval-2018 Task 10: A

Support-Vector-Machine-Based Method for Detecting Semantic Difference using Taxonomy and Word Embedding Features

Sunny Lai, Kwong Sak Leung, and Yee Leung

14:30–14:45 SemEval-2018 Task 11: Machine Comprehension Using Commonsense Knowledge

Simon Ostermann, Michael Roth, Ashutosh Modi, Stefan Thater, and Manfred Pinkal

14:45–15:00 Yuanfudao at SemEval-2018 Task 11: Three-way Attention and Relational Knowledge for Commonsense Machine Comprehension

Liang Wang, Meng Sun, Wei Zhao, Kewei Shen, and Jingming Liu

15:00–15:15 SemEval-2018 Task 12: The Argument Reasoning Comprehension Task

Ivan Habernal, Henning Wachsmuth, Iryna Gurevych, and Benno Stein

15:15–15:30 GIST at SemEval-2018 Task 12: A network transferring inference knowledge to Argument Reasoning Comprehension task

HongSeok Choi and Hyunjoo Lee

15:30–16:00 Coffee

16:00–16:30 **Discussion**

16:30–17:30 **Poster Session**

- 16:30–17:30 LightRel at SemEval-2018 Task 7: Lightweight and Fast Relation Classification
Tyler Renslow and Günter Neumann
- 16:30–17:30 OhioState at SemEval-2018 Task 7: Exploiting Data Augmentation for Relation Classification in Scientific Papers Using Piecewise Convolutional Neural Networks
Dushyanta Dhyani
- 16:30–17:30 The UWNLP system at SemEval-2018 Task 7: Neural Relation Extraction Model with Selectively Incorporated Concept Embeddings
Yi Luan, Mari Ostendorf, and Hannaneh Hajishirzi
- 16:30–17:30 UC3M-NII Team at SemEval-2018 Task 7: Semantic Relation Classification in Scientific Papers via Convolutional Neural Network
Víctor Suárez-Paniagua, Isabel Segura-Bedmar, and Akiko Aizawa
- 16:30–17:30 MIT-MEDG at SemEval-2018 Task 7: Semantic Relation Classification via Convolution Neural Network
Di Jin, Franck Dernoncourt, Elena Sergeeva, Matthew McDermott, and Geetika Chauhan
- 16:30–17:30 SIRIUS-LTG-UiO at SemEval-2018 Task 7: Convolutional Neural Networks with Shortest Dependency Paths for Semantic Relation Extraction and Classification in Scientific Papers
Farhad Nooralahzadeh, Lilja Øvrelid, and Jan Tore Lønning
- 16:30–17:30 IRCMS at SemEval-2018 Task 7 : Evaluating a basic CNN Method and Traditional Pipeline Method for Relation Classification
Zhongbo Yin, Zhunchen Luo, Luo Wei, Mao Bin, Tian Changhai, Ye Yuming, and Wu Shuai
- 16:30–17:30 BF3R at SemEval-2018 Task 7: Evaluating Two Relation Extraction Tools for Finding Semantic Relations in Biomedical Abstracts
Mariana Neves, Daniel Butzke, Gilbert Schönfelder, and Barbara Grune
- 16:30–17:30 Texterra at SemEval-2018 Task 7: Exploiting Syntactic Information for Relation Extraction and Classification in Scientific Papers
Andrey Sysoev and Vladimir Mayorov
- 16:30–17:30 UniMa at SemEval-2018 Task 7: Semantic Relation Extraction and Classification from Scientific Publications
Thorsten Keiper, Zhonghao Lyu, Sara Pooladzadeh, Yuan Xu, Jingyi Zhang, Anne Lauscher, and Simone Paolo Ponzetto
- 16:30–17:30 GU IRLAB at SemEval-2018 Task 7: Tree-LSTMs for Scientific Relation Classification
Sean MacAvaney, Luca Soldaini, Arman Cohan, and Nazli Goharian
- 16:30–17:30 ClaiRE at SemEval-2018 Task 7: Classification of Relations using Embeddings
Lena Hettinger, Alexander Dallmann, Albin Zehe, Thomas Niebler, and Andreas Hotho
- 16:30–17:30 TakeLab at SemEval-2018 Task 7: Combining Sparse and Dense Features for Relation Classification in Scientific Texts
Martin Gluhak, Maria Pia di Buono, Abbas Akkasi, and Jan Šnajder
- 16:30–17:30 NEUROSENT-PDI at SemEval-2018 Task 7: Discovering Textual Relations With a Neural Network Model
Mauro Dragoni
- 16:30–17:30 SciREL at SemEval-2018 Task 7: A System for Semantic Relation Extraction and Classification
Darshini Mahendran, Chathurika Brahmana, and Bridget McInnes

- 16:30–17:30 NTNU at SemEval-2018 Task 7: Classifier Ensembling for Semantic Relation Identification and Classification in Scientific Papers
Biswanath Barik, Utpal Kumar Sikdar, and Björn Gambäck
- 16:30–17:30 Talla at SemEval-2018 Task 7: Hybrid Loss Optimization for Relation Classification using Convolutional Neural Networks
Bhanu Pratap, Daniel Shank, Oladipo Ositelu, and Byron Galbraith
- 16:30–17:30 TeamDL at SemEval-2018 Task 8: Cybersecurity Text Analysis using Convolutional Neural Network and Conditional Random Fields
Manikandan R, Krishna Madgula, and Snehanshu Saha
- 16:30–17:30 HCCL at SemEval-2018 Task 8: An End-to-End System for Sequence Labeling from Cybersecurity Reports
Mingming Fu, Xuemin Zhao, and Yonghong Yan
- 16:30–17:30 UMBC at SemEval-2018 Task 8: Understanding Text about Malware
Ankur Padia, Arpita Roy, Taneeya Satyapanich, Francis Ferraro, Shimei Pan, Youngja Park, Anupam Joshi, and Tim Finin
- 16:30–17:30 Villani at SemEval-2018 Task 8: Semantic Extraction from Cybersecurity Reports using Representation Learning
Pablo Loyola, Kugamoorthy Gajananan, Yuji Watanabe, and Fumiko Satoh
- 16:30–17:30 Flytxt_NTNU at SemEval-2018 Task 8: Identifying and Classifying Malware Text Using Conditional Random Fields and Naïve Bayes Classifiers
Utpal Kumar Sikdar, Biswanath Barik, and Björn Gambäck
- 16:30–17:30 Digital Operatives at SemEval-2018 Task 8: Using dependency features for malware NLP
Chris Brew
- 16:30–17:30 Apollo at SemEval-2018 Task 9: Detecting Hypernymy Relations Using Syntactic Dependencies
Mihaela Onofrei, Ionut Hulub, Diana Trandabat, and Daniela Gifu
- 16:30–17:30 SJTU-NLP at SemEval-2018 Task 9: Neural Hypernym Discovery with Term Embeddings
Zhousheng Zhang, Jiangtong Li, Hai Zhao, and Bingjie Tang
- 16:30–17:30 NLP_HZ at SemEval-2018 Task 9: a Nearest Neighbor Approach
Wei Qiu, Mosha Chen, Linlin Li, and Luo Si
- 16:30–17:30 UMDuluth-CS8761 at SemEval-2018 Task9: Hypernym Discovery using Hearst Patterns, Co-occurrence frequencies and Word Embeddings
Arshia Zernab Hassan, Manikya Swathi Vallabhajosyula, and Ted Pedersen
- 16:30–17:30 EXPR at SemEval-2018 Task 9: A Combined Approach for Hypernym Discovery
Ahmad Issa Alaa Aldine, Mounira Harzallah, Giuseppe Berio, Nicolas Béchet, and Ahmad Faour
- 16:30–17:30 ADAPT at SemEval-2018 Task 9: Skip-Gram Word Embeddings for Unsupervised Hypernym Discovery in Specialised Corpora
Alfredo Maldonado and Filip Klubička
- 16:30–17:30 300-sparsans at SemEval-2018 Task 9: Hypernymy as interaction of sparse attributes
Gábor Berend, Márton Makrai, and Péter Földiák
- 16:30–17:30 UWB at SemEval-2018 Task 10: Capturing Discriminative Attributes from Word Distributions
Tomáš Brychcín, Tomáš Hercig, Josef Steinberger, and Michal Konkol
- 16:30–17:30 Meaning_space at SemEval-2018 Task 10: Combining explicitly encoded knowledge with information extracted from word embeddings
Pia Sommerauer, Antske Fokkens, and Piek Vossen

- 16:30–17:30 GHH at SemEval-2018 Task 10: Discovering Discriminative Attributes in Distributional Semantics
Mohammed Attia, Younes Samih, Manaal Faruqui, and Wolfgang Maier
- 16:30–17:30 CitiusNLP at SemEval-2018 Task 10: The Use of Transparent Distributional Models and Salient Contexts to Discriminate Word Attributes
Pablo Gamallo
- 16:30–17:30 THU_NGN at SemEval-2018 Task 10: Capturing Discriminative Attributes with MLP-CNN model
Chuhan Wu, Fangzhao Wu, Sixing Wu, Zhigang Yuan, and Yongfeng Huang
- 16:30–17:30 ALB at SemEval-2018 Task 10: A System for Capturing Discriminative Attributes
Bogdan Dumitru, Alina Maria Ciobanu, and Liviu P. Dinu
- 16:30–17:30 ELiRF-UPV at SemEval-2018 Task 10: Capturing Discriminative Attributes with Knowledge Graphs and Wikipedia
José-Ángel González, Lluís-F. Hurtado, Encarna Segarra, and Ferran Pla
- 16:30–17:30 Wolves at SemEval-2018 Task 10: Semantic Discrimination based on Knowledge and Association
Shiva Taslimipoor, Omid Rohanian, Le An Ha, Gloria Corpas Pastor, and Ruslan Mitkov
- 16:30–17:30 UNAM at SemEval-2018 Task 10: Unsupervised Semantic Discriminative Attribute Identification in Neural Word Embedding Cones
Ignacio Arroyo-Fernández, Ivan Vladimir Meza Ruiz, and Carlos-Francisco Meéndez-Cruz
- 16:30–17:30 Luminoso at SemEval-2018 Task 10: Distinguishing Attributes Using Text Corpora and Relational Knowledge
Robert Speer and Joanna Lowry-Duda
- 16:30–17:30 BomJi at SemEval-2018 Task 10: Combining Vector-, Pattern- and Graph-based Information to Identify Discriminative Attributes
Enrico Santus, Chris Biemann, and Emmanuele Chersoni
- 16:30–17:30 Igevorse at SemEval-2018 Task 10: Exploring an Impact of Word Embeddings Concatenation for Capturing Discriminative Attributes
Maxim Grishin
- 16:30–17:30 ECNU at SemEval-2018 Task 10: Evaluating Simple but Effective Features on Machine Learning Methods for Semantic Difference Detection
Yunxiao Zhou, Man Lan, and Yuanbin Wu
- 16:30–17:30 AmritaNLP at SemEval-2018 Task 10: Capturing discriminative attributes using convolution neural network over global vector representation.
Vivek Vinayan, Anand Kumar M, and Soman K P
- 16:30–17:30 Discriminator at SemEval-2018 Task 10: Minimally Supervised Discrimination
Artur Kulmizev, Mostafa Abdou, Vinit Ravishankar, and Malvina Nissim
- 16:30–17:30 UNBNLP at SemEval-2018 Task 10: Evaluating unsupervised approaches to capturing discriminative attributes
Milton King, Ali Hakimi Parizi, and Paul Cook
- 16:30–17:30 ABDN at SemEval-2018 Task 10: Recognising Discriminative Attributes using Context Embeddings and WordNet
Rui Mao, Guanyi Chen, Ruijie Li, and Chenghua Lin
- 16:30–17:30 UMD at SemEval-2018 Task 10: Can Word Embeddings Capture Discriminative Attributes?
Alexander Zhang and Marine Carpuat
- 16:30–17:30 NTU NLP Lab System at SemEval-2018 Task 10: Verifying Semantic Differences by Integrating Distributional Information and Expert Knowledge
Yow-Ting Shieue, Hen-Hsien Huang, and Hsin-Hsi Chen
-

- 16:30–17:30 ELiRF-UPV at SemEval-2018 Task 11: Machine Comprehension using Commonsense Knowledge
José-Ángel González, Lluís-F. Hurtado, Encarna Segarra, and Ferran Pla
- 16:30–17:30 YNU_AI1799 at SemEval-2018 Task 11: Machine Comprehension using Commonsense Knowledge of Different model ensemble
Liu Qingxun, Yao Hongdou, Zhou Xiaobing, and Xie Ge
- 16:30–17:30 YNU_Deep at SemEval-2018 Task 11: An Ensemble of Attention-based BiLSTM Models for Machine Comprehension
Peng Ding and Xiaobing Zhou
- 16:30–17:30 ECNU at SemEval-2018 Task 11: Using Deep Learning Method to Address Machine Comprehension Task
Yixuan Sheng, Man Lan, and Yuanbin Wu
- 16:30–17:30 CSReader at SemEval-2018 Task 11: Multiple Choice Question Answering as Textual Entailment
Zhengping Jiang and Qi Sun
- 16:30–17:30 YNU-HPCC at Semeval-2018 Task 11: Using an Attention-based CNN-LSTM for Machine Comprehension using Commonsense Knowledge
Hang Yuan, Jin Wang, and Xuejie Zhang
- 16:30–17:30 Jiangnan at SemEval-2018 Task 11: Deep Neural Network with Attention Method for Machine Comprehension Task
Jiangnan Xia
- 16:30–17:30 IUCM at SemEval-2018 Task 11: Similar-Topic Texts as a Comprehension Knowledge Source
Sofia Reznikova and Leon Derczynski
- 16:30–17:30 Lyb3b at SemEval-2018 Task 11: Machine Comprehension Task using Deep Learning Models
Yongbin Li and Xiaobing Zhou
- 16:30–17:30 MITRE at SemEval-2018 Task 11: Commonsense Reasoning without Commonsense Knowledge
Elizabeth Merkhofer, John Henderson, David Bloom, Laura Strickhart, and Guido Zarrella
- 16:30–17:30 SNU_IDS at SemEval-2018 Task 12: Sentence Encoder with Contextualized Vectors for Argument Reasoning Comprehension
Taeuk Kim, Jihun Choi, and Sang-goo Lee
- 16:30–17:30 ITNLP-ARC at SemEval-2018 Task 12: Argument Reasoning Comprehension with Attention
Wenjie Liu, Chengjie Sun, Lei Lin, and Bingquan Liu
- 16:30–17:30 ECNU at SemEval-2018 Task 12: An End-to-End Attention-based Neural Network for the Argument Reasoning Comprehension Task
Junfeng Tian, Man Lan, and Yuanbin Wu
- 16:30–17:30 NLITrans at SemEval-2018 Task 12: Transfer of Semantic Knowledge for Argument Comprehension
Timothy Niven and Hung-Yu Kao
- 16:30–17:30 BLCU_NLP at SemEval-2018 Task 12: An Ensemble Model for Argument Reasoning Based on Hierarchical Attention
Meiqian Zhao, Chunhua Liu, Lu Liu, Yan Zhao, and Dong Yu
- 16:30–17:30 YNU-HPCC at SemEval-2018 Task 12: The Argument Reasoning Comprehension Task Using a Bi-directional LSTM with Attention Model
Quanlei Liao, Xutao Yang, Jin Wang, and Xuejie Zhang
- 16:30–17:30 HHU at SemEval-2018 Task 12: Analyzing an Ensemble-based Deep Learning Approach for the Argument Mining Task of Choosing the Correct Warrant
Matthias Liebeck, Andreas Funke, and Stefan Conrad

- 16:30–17:30 YNU Deep at SemEval-2018 Task 12: A BiLSTM Model with Neural Attention for Argument Reasoning Comprehension
Peng Ding and Xiaobing Zhou
- 16:30–17:30 UniMelb at SemEval-2018 Task 12: Generative Implication using LSTMs, Siamese Networks and Semantic Representations with Synonym Fuzzing
Anirudh Joshi, Timothy Baldwin, Richard O. Sinnott, and Cecile Paris
- 16:30–17:30 Joker at SemEval-2018 Task 12: The Argument Reasoning Comprehension with Neural Attention
Sui Guobin, Chao Wenhan, and Luo Zhunchen
- 16:30–17:30 TakeLab at SemEval-2018 Task12: Argument Reasoning Comprehension with Skip-Thought Vectors
Ana Brassard, Tin Kuculo, Filip Boltuzic, and Jan Šnajder
- 16:30–17:30 Lyb3b at SemEval-2018 Task 12: Ensemble-based Deep Learning Models for Argument Reasoning Comprehension Task
Yongbin Li and Xiaobing Zhou
- 16:30–17:30 TRANSRW at SemEval-2018 Task 12: Transforming Semantic Representations for Argument Reasoning Comprehension
Zhimin Chen, Wei Song, and Lizhen Liu

*SEM: Seventh Joint Conference On Lexical And Computational Semantics

Organizers: *Malvina Nissim, Jonathan Berant, and Alessandro Lenci*

Venue: Strand 12B

Tuesday, June 5, 2018

08:00–17:00 **NAACL Registration**

09:00–10:30 **Session 1**

09:15–09:30 **Opening Remarks**

09:30–10:30 **Invited Talk by Ellie Pavlick (Brown University): Why Should we Care about Linguistics?**

10:30–11:00 **Coffee Break**

11:00–12:30 **Session 2**

11:00–11:30 Resolving Event Coreference with Supervised Representation Learning and Clustering-Oriented Regularization

Kian Kenyon-Dean, Jackie Chi Kit Cheung, and Doina Precup

11:30–12:00 Learning distributed event representations with a multi-task approach

Xudong Hong, Asad Sayeed, and Vera Demberg

12:00–12:15 Assessing Meaning Components in German Complex Verbs: A Collection of Source-Target Domains and Directionality

Sabine Schulte im Walde, Maximilian Köper, and Sylvia Springorum

12:15–12:30 Learning Neural Word Salience Scores

Krasen Samardzhiev, Andrew Gargett, and Danushka Bollegala

12:30–14:00 **Lunch Break**

14:00–15:30 **Session 3**

14:00–14:30 Examining Gender and Race Bias in Two Hundred Sentiment Analysis Systems

Svetlana Kiritchenko and Saif Mohammad

14:30–15:00 Graph Algebraic Combinatory Categorial Grammar

Sebastian Beschke and Wolfgang Menzel

15:00–15:15 Mixing Context Granularities for Improved Entity Linking on Question Answering Data across Entity Categories

Daniil Sorokin and Iryna Gurevych

15:15–15:30 Quantitative Semantic Variation in the Contexts of Concrete and Abstract Words

Daniela Naumann, Diego Frassinelli, and Sabine Schulte im Walde

15:30–16:00 **Coffee Break**

16:00–18:00 **Session 4**

16:00–16:50 **Poster Booster**

16:50–18:00 **Poster Session**

- EmoWordNet: Automatic Expansion of Emotion Lexicon Using English WordNet
Gilbert Badaro, Hussein Jundi, Hazem Hajj, and Wassim El-Hajj
- The Limitations of Cross-language Word Embeddings Evaluation
Amir Bakarov, Roman Suvorov, and Ilya Sochenkov
- How Gender and Skin Tone Modifiers Affect Emoji Semantics in Twitter
Francesco Barbieri and Jose Camacho-Collados
- Element-wise Bilinear Interaction for Sentence Matching
Jihun Choi, Taeuk Kim, and Sang-goo Lee
- Named Graphs for Semantic Representation
Richard Crouch and Aikaterini-Lida Kalouli
- Learning Patient Representations from Text
Dmitriy Dligach and Timothy Miller
- Polarity Computations in Flexible Categorial Grammar
Hai Hu and Larry Moss
- Coarse Lexical Frame Acquisition at the Syntax—Semantics Interface Using a Latent-Variable PCFG Model
Laura Kallmeyer, Behrang QasemiZadeh, and Jackie Chi Kit Cheung
- Halo: Learning Semantics-Aware Representations for Cross-Lingual Information Extraction
Hongyuan Mei, Sheng Zhang, Kevin Duh, and Benjamin Van Durme
- Exploiting Partially Annotated Data in Temporal Relation Extraction
Qiang Ning, Zhongzhi Yu, Chuchu Fan, and Dan Roth
- Predicting Word Embeddings Variability
Benedicte Pierrejean and Ludovic Tanguy
- Integrating Multiplicative Features into Supervised Distributional Methods for Lexical Entailment
Tu Vu and Vered Shwartz
- Deep Affix Features Improve Neural Named Entity Recognizers
Vikas Yadav, Rebecca Sharp, and Steven Bethard
- Fine-grained Entity Typing through Increased Discourse Context and Adaptive Classification Thresholds
Sheng Zhang, Kevin Duh, and Benjamin Van Durme

Wednesday, June 6, 2018

08:00–16:00 NAACL Registration

09:00–10:30 Session 5

09:00–10:00 Invited Talk by Christopher Potts (Stanford University): Linguists for Deep Learning; or How I Learned to Stop Worrying and Love Neural Networks

10:00–10:30 Hypothesis Only Baselines in Natural Language Inference
Adam Poltak, Jason Naradowsky, Aparajita Haldar, Rachel Rudinger, and Benjamin Van Durme

10:30–11:00 Coffee Break

11:00–12:15 Session 6

11:00–11:30 Quality Signals in Generated Stories
Manasvi Sagarkar, John Wieting, Lifu Tu, and Kevin Gimpel

11:30–12:00 Term Definitions Help Hypernymy Detection
Wenpeng Yin and Dan Roth

12:00–12:15 Agree or Disagree: Predicting Judgments on Nuanced Assertions
Michael Wojatzki, Torsten Zesch, Saif Mohammad, and Svetlana Kiritchenko

12:15–14:00 Lunch Break

14:00–15:30 Session 7

14:00–14:30 A Multimodal Translation-Based Approach for Knowledge Graph Representation Learning
Hatem Mousselly Sergieh, Teresa Botschen, Iryna Gurevych, and Stefan Roth

14:30–15:30 Putting Semantics into Semantic Roles
James Allen and Choh Man Teng

15:00–15:30 Measuring Frame Instance Relatedness
Valerio Basile, Roque Lopez Condori, and Elena Cabrio

15:30–16:00 Coffee Break

16:00–16:30 Solving Feature Sparseness in Text Classification using Core-Periphery Decomposition
Xia Cui, Sadamori Kojaku, Naoki Masuda, and Danushka Bollegala

16:30–17:00 Robust Handling of Polysemy via Sparse Representations
Abhijit Mahabal, Dan Roth, and Sid Mittal

17:00–17:30 Multiplicative Tree-Structured Long Short-Term Memory Networks for Semantic Representations
Nam Khanh Tran and Weiwei Cheng

GenDeep: Generalization in the Age of Deep Learning

Organizers: *Yonatan Bisk, Omer Levy, and Mark Yatskar*

Venue: Strand 11

Tuesday, June 5, 2018

08:00–17:00 **NAACL Registration**

09:00–09:15 **Welcome**

09:15–09:50 **Yejin Choi**

09:50–10:25 **Dan Roth**

10:25–10:35 **Break**

10:35–11:10 **Percy Liang**

11:10–11:45 **Ndapa Nakashole**

11:45–12:20 **Hal Daumé III**

12:20–13:30 **Lunch**

13:30–14:30 **Poster Session**

- Towards Inference-Oriented Reading Comprehension: ParallelQA
Soumya Wadhwa, Varsha Embar, Matthias Grabmair, and Eric Nyberg
- Commonsense mining as knowledge base completion? A study on the impact of novelty
Stanislaw Jastrzebski, Dzmitry Bahdanau, Seyedarian Hosseini, Michael Noukhovitch, Yoshua Bengio, and Jackie Chi Kit Cheung
- Deep learning evaluation using deep linguistic processing
Alexander Kuhnle and Ann Copestake
- The Fine Line between Linguistic Generalization and Failure in Seq2Seq-Attention Models
Noah Weber, Leena Shekhar, and Niranjan Balasubramanian
- Extrapolation in NLP
Jeff Mitchell, Pontus Stenetorp, Pasquale Minervini, and Sebastian Riedel

14:45–15:20 **Sam Bowman**

15:20–15:55 **Devi Parikh**

15:55–16:10 **Break**

16:10–17:10 **Panel**

17:10–17:15 **Closing**

BEA: 13th Workshop on Innovative Use of NLP for Building Educational Applications

Organizers: *Joel Tetreault, Jill Burstein, Ekaterina Kochmar, Claudia Leacock, and Helen Yannakoudakis*

Venue: Strand 12A

Tuesday, June 5, 2018

08:00–17:00 NAACL Registration

08:30–09:00 Loading of Oral Presentations

09:00–10:30 Oral Presentations (Speech, Dialogue & Reading)

09:00–09:15 Opening Remarks

09:15–09:40 Using exemplar responses for training and evaluating automated speech scoring systems

Anastassia Loukina, Klaus Zechner, James Bruno, and Beata Beigman Klebanov

09:40–10:05 Using Paraphrasing and Memory-Augmented Models to Combat Data Sparsity in Question Interpretation with a Virtual Patient Dialogue System

Lifeng Jin, David King, Amad Hussein, Michael White, and Douglas Danforth

10:05–10:30 Predicting misreadings from gaze in children with reading difficulties

Joachim Bingel, Maria Barrett, and Sigrid Klerke

10:30–11:00 Mid-morning coffee/decaf/tea refresh

11:00–12:30 Oral Presentations (Passage Selection, Text Complexity & Reading; Shared Task Reports)

11:00–11:25 Automatic Input Enrichment for Selecting Reading Material: An Online Study with English Teachers

Maria Chinkina, Ankita Oswal, and Detmar Meurers

11:25–11:50 Estimating Linguistic Complexity for Science Texts

Farah Nadeem and Mari Ostendorf

11:50–12:10 Second Language Acquisition Modeling

Burr Settles, Chris Brust, Erin Gustafson, Masato Hagiwara, and Nitin Madnani

12:10–12:30 A Report on the Complex Word Identification Shared Task 2018

Seid Muhie Yimam, Chris Biemann, Shervin Malmasi, Gustavo Paetzold, Lucia Specia, Sanja Štajner, Anaïs Tack, and Marcos Zampieri

12:30–14:00 Lunch

14:00–15:30 BEA & Shared Task Poster and Demo Session

14:00–14:45 Poster Session A

14:00–14:45 BEA papers

14:00–14:45 Towards Single Word Lexical Complexity Prediction

David Alfter and Elena Volodina

- 14:00–14:45 COAST - Customizable Online Syllable Enhancement in Texts. A flexible framework for automatically enhancing reading materials
Heiko Holz, Zarah Weiss, Oliver Brehm, and Detmar Meurers
- 14:00–14:45 Annotating picture description task responses for content analysis
Levi King and Markus Dickinson
- 14:00–14:45 Annotating Student Talk in Text-based Classroom Discussions
Luca Lugini, Diane Litman, Amanda Godley, and Christopher Olshefski
- 14:00–14:45 Toward Automatically Measuring Learner Ability from Human-Machine Dialog Interactions using Novel Psychometric Models
Vikram Ramanarayanan and Michelle LaMar
- 14:00–14:45 Generating Feedback for English Foreign Language Exercises
Björn Rudzewitz, Ramon Ziai, Kordula De Kuthy, Verena Möller, Florian Nuxoll, and Detmar Meurers
- 14:00–14:45 NT2Lex: A CEFR-Graded Lexical Resource for Dutch as a Foreign Language Linked to Open Dutch WordNet
Anaïs Tack, Thomas François, Piet Desmet, and Cédrick Fairon
- 14:00–14:45 Experiments with Universal CEFR Classification
Sowmya Vajjala and Taraka Rama
- 14:00–14:45 Chengyu Cloze Test
Zhiying Jiang, Boliang Zhang, Lifu Huang, and Heng Ji
- 14:00–14:45 CWI Shared Task papers**
- 14:00–14:45 LaSTUS/TALN at Complex Word Identification (CWI) 2018 Shared Task
Ahmed AbuRa'ed and Horacio Saggion
- 14:00–14:45 Cross-lingual complex word identification with multitask learning
Joachim Bingel and Johannes Bjerva
- 14:00–14:45 UnibucKernel: A kernel-based learning method for complex word identification
Andrei Butnaru and Radu Tudor Ionescu
- 14:00–14:45 CAMB at CWI Shared Task 2018: Complex Word Identification with Ensemble-Based Voting
Sian Gooding and Ekaterina Kochmar
- 14:00–14:45 Complex Word Identification Based on Frequency in a Learner Corpus
Tomoyuki Kajiwara and Mamoru Komachi
- 14:00–14:45 The Whole is Greater than the Sum of its Parts: Towards the Effectiveness of Voting Ensemble Classifiers for Complex Word Identification
Nikhil Wani, Sandeep Mathias, Jayashree Aanand Gajjam, and Pushpak Bhattacharya
- 14:00–14:45 SLAM Shared Task papers**
- 14:00–14:45 GrotocoSLAM: Second Language Acquisition Modeling with Simple Features, Learners and Task-wise Models
Sigrid Klerke, Héctor Martínez Alonso, and Barbara Plank
- 14:00–14:45 Context Based Approach for Second Language Acquisition
Nihal V. Nayak and Arjun R. Rao
- 14:00–14:45 Second Language Acquisition Modeling: An Ensemble Approach
Anton Osika, Susanna Nilsson, Andrii Sydorchuk, Faruk Sahin, and Anders Huss
- 14:00–14:45 Modeling Second-Language Learning from a Psychological Perspective
Alexander Rich, Pamela Osborn Popp, David Halpern, Anselm Rothe, and Todd Gureckis
- 14:00–14:45 A Memory-Sensitive Classification Model of Errors in Early Second Language Learning
Brendan Tomoschuk and Jarrett Lovelett
- 14:45–15:30 **Poster Session B**
-

14:45–15:30 **BEA papers**

- 14:45–15:30 Annotation and Classification of Sentence-level Revision Improvement
Tazin Afrin and Diane Litman
- 14:45–15:30 Language Model Based Grammatical Error Correction without Annotated Training Data
Christopher Bryant and Ted Briscoe
- 14:45–15:30 A Semantic Role-based Approach to Open-Domain Automatic Question Generation
Michael Flor and Brian Riordan
- 14:45–15:30 Automated Content Analysis: A Case Study of Computer Science Student Summaries
Yanjun Gao, Patricia M.Davies, and Rebecca J. Passonneau
- 14:45–15:30 Toward Data-Driven Tutorial Question Answering with Deep Learning Conversational Models
Mayank Kulkarni and Kristy Boyer
- 14:45–15:30 Distractor Generation for Multiple Choice Questions Using Learning to Rank
Chen Liang, Xiao Yang, Neisarg Dave, Drew Wham, Bart Pursel, and C Lee Giles
- 14:45–15:30 A Portuguese Native Language Identification Dataset
Iria del Río Gayo, Marcos Zampieri, and Shervin Malmasi
- 14:45–15:30 OneStopEnglish corpus: A new corpus for automatic readability assessment and text simplification
Sowmya Vajjala and Ivana Lucic
- 14:45–15:30 The Effect of Adding Authorship Knowledge in Automated Text Scoring
Meng Zhang, Xie Chen, Ronan Cummins, Øistein E. Andersen, and Ted Briscoe

14:45–15:30 **CWI Shared Task papers**

- 14:45–15:30 SBGU at the Complex Word Identification 2018 Shared Task
David Alfter and Ildikó Pilán
- 14:45–15:30 Complex Word Identification: Convolutional Neural Network vs. Feature Engineering
Segun Taofeek Aroyehun, Jason Angel, Daniel Alejandro Pérez Alvarez, and Alexander Gelbukh
- 14:45–15:30 Deep Learning Architecture for Complex Word Identification
Dirk De Hertog and Anaïs Tack
- 14:45–15:30 NILC at CWI 2018: Exploring Feature Engineering and Feature Learning
Nathan Hartmann and Leandro Borges dos Santos
- 14:45–15:30 Complex Word Identification Using Character n-grams
Maja Popović

14:45–15:30 **SLAM Shared Task papers**

- 14:45–15:30 Predicting Second Language Learner Successes and Mistakes by Means of Conjunctive Features
Yves Bestgen
- 14:45–15:30 Feature Engineering for Second Language Acquisition Modeling
Guanliang Chen, Claudia Hauff, and Geert-Jan Houben
- 14:45–15:30 TMU System for SLAM-2018
Masahiro Kaneko, Tomoyuki Kajiwara, and Mamoru Komachi
- 14:45–15:30 Deep Factorization Machines for Knowledge Tracing
Jill-Jenn Vie
- 14:45–15:30 CLUF: a Neural Model for Second Language Acquisition Modeling
Shuyao Xu, Jin Chen, and Long Qin
- 14:45–15:30 Neural sequence modelling for learner error prediction
Zheng Yuan

15:30–16:00 **Mid-afternoon Snacks**

16:00–17:30 **Oral Presentations (Item Generation, Essay/Content Scoring, & Writing)**

16:00–16:25 Automatic Distractor Suggestion for Multiple-Choice Tests Using Concept Embeddings and Information Retrieval
Le An Ha and Victoria Yaneva

16:25–16:50 Co-Attention Based Neural Network for Source-Dependent Essay Scoring
Haoran Zhang and Diane Litman

16:50–17:15 Cross-Lingual Content Scoring
Andrea Horbach, Sebastian Stemmmanns, and Torsten Zesch

17:15–17:30 **Closing Remarks**

Ethics-NLP: Second ACL Workshop on Ethics in Natural Language Processing

Organizers: *Mark Alfano, Dirk Hovy, Margaret Mitchell, and Michael Strube*

Venue: Bolden 1

Tuesday, June 5, 2018

08:00–17:00 **NAACL Registration**

09:00–10:30 **Session 1**

09:00–09:15 **Welcome**

09:15–09:40 On the Utility of Lay Summaries and AI Safety Disclosures: Toward Robust, Open Research Oversight
Allen Schmalz

09:40–10:05 #MeToo Alexa: How Conversational Systems Respond to Sexual Harassment
Amanda Cercas Curry and Verena Rieser

10:05–10:30 **Invited Talk: Examining Gender and Race Bias in Two Hundred Sentiment Analysis Systems (Svetlana Kiritchenko and Saif Mohammad)**

10:30–11:00 **Coffee**

11:00–12:30 **Session 2**

11:00–11:45 **Invited Talk**

11:45–12:30 **Invited Talk**

12:30–14:00 **Lunch**

14:00–15:30 **Session 3**

14:00–14:45 **Invited Talk**

14:45–15:30 **Invited Talk**

15:30–16:00 **Coffee Break**

16:00–17:00 **Science cafe roundtable discussions**

17:00–17:15 **Reaction to roundtable**

17:15–18:00 **Invited talk**

CLPsych: Fifth Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality

Organizers: *Kate Loveys, Kate Niederhoffer, Emily Prud'hommeaux, Rebecca Resnik, and Philip Resnik*

Venue: Bolden 5

Tuesday, June 5, 2018

08:00–17:00 **NAACL Registration**

09:00–09:15 **Opening Remarks**

09:15–10:35 **Workshop Session I: Presentations with Discussant Commentary**

- What type of happiness are you looking for? - A closer look at detecting mental health from language
Alina Arseniev-Koehler, Sharon Mozgai, and Stefan Scherer
- A Linguistically-Informed Fusion Approach for Multimodal Depression Detection
Michelle Morales, Stefan Scherer, and Rivka Levitan
- Expert, Crowdsourced, and Machine Assessment of Suicide Risk via Online Postings
Han-Chin Shing, Suraj Nair, Ayah Zirikly, Meir Friedenberg, Hal Daumé III, and Philip Resnik

10:35–10:55 **Break**

10:55–11:40 **Plenary Session**

11:40–12:40 **Workshop Session II: Shared Task Presentations with Discussant Commentary**

- CLPsych 2018 Shared Task: Predicting Current and Future Psychological Health from Childhood Essays
Veronica Lynn, Alissa Goodman, Kate Niederhoffer, Kate Loveys, Philip Resnik, and H. Andrew Schwartz
- An Approach to the CLPsych 2018 Shared Task Using Top-Down Text Representation and Simple Bottom-Up Model Selection
Micah Iserman, Molly Ireland, Andrew Littlefield, Tyler Davis, and Sage Maliepaard

12:40–14:00 **Lunch and Poster Session**

- Using contextual information for automatic triage of posts in a peer-support forum
Edgar Altszyler, Ariel J. Berenstein, David Milne, Rafael A. Calvo, and Diego Fernandez Slezak
- Hierarchical neural model with attention mechanisms for the classification of social media text related to mental health
Julia Iye, George Gkotsis, Rina Dutta, Robert Stewart, and Sumithra Velupillai

- Cross-cultural differences in language markers of depression online
Kate Loveys, Jonathan Torrez, Alex Fine, Glen Moriarty, and Glen Coppersmith
- Deep Learning for Depression Detection of Twitter Users
Ahmed Husseini Orabi, Prasadith Buddhitha, Mahmoud Husseini Orabi, and Diana Inkpen
- Current and Future Psychological Health Prediction using Language and Socio-Demographics of Children for the CLPsych 2018 Shared Task
Sharath Chandra Guntuku, Salvatore Giorgi, and Lyle Ungar
- Predicting Psychological Health from Childhood Essays with Convolutional Neural Networks for the CLPsych 2018 Shared Task (Team UKNLP)
Anthony Rios, Tung Tran, and Ramakanth Kavuluru
- A Psychologically Informed Approach to CLPsych Shared Task 2018
Almog Simchon and Michael Gilead
- Predicting Psychological Health from Childhood Essays. The UGent-IDLab CLPsych 2018 Shared Task System.
Klim Zaporojets, Lucas Sterckx, Johannes Deleu, Thomas Demeester, and Chris Develder
- Can adult mental health be predicted by childhood future-self narratives? Insights from the CLPsych 2018 Shared Task
Kylie Radford, Louise Lavrencic, Ruth Peters, Kim Kiely, Ben Hachey, Scott Nowson, and Will Radford

14:00–15:00 **Workshop Session III: Presentations with Discussant Commentary**

- Automatic Detection of Incoherent Speech for Diagnosing Schizophrenia
Dan Iter, Jong Yoon, and Dan Jurafsky
- Oral-Motor and Lexical Diversity During Naturalistic Conversations in Adults with Autism Spectrum Disorder
Julia Parish-Morris, Evangelos Sariyanidi, Casey Zampella, G. Keith Bartley, Emily Ferguson, Ashley A. Pallathra, Leila Bateman, Samantha Plate, Meredith Cola, Juhi Pandey, Edward S. Brodkin, Robert T. Schultz, and Birkana Tunc

15:00–15:30 **Workshop Session IV: Short Presentations**

- Dynamics of an idiosyncrasy of a Russian suicidal blogger
Tatiana Litvinova, Olga Litvinova, and Pavel Seredin
- RSDD-Time: Temporal Annotation of Self-Reported Mental Health Diagnoses
Sean MacAvaney, Bart Desmet, Arman Cohan, Luca Soldaini, Andrew Yates, Ayah Zirikly, and Nazli Goharian

15:30–16:00 **Break**

16:00–16:45 **Workshop Session V: Short Presentations**

- Predicting Human Trustworthiness from Facebook Language
Mohammadzaman Zamani, Anneke Buffone, and H. Andrew Schwartz
- Within and Between-Person Differences in Language Used Across Anxiety Support and Neutral Reddit Communities
Molly Ireland and Micah Iserman
- Helping or Hurting? Predicting Changes in Users' Risk of Self-Harm Through Online Community Interactions
Luca Soldaini, Timothy Walsh, Arman Cohan, Julien Han, and Nazli Goharian

16:45–17:30 **Workshop General Discussion and Closing Remarks**

SemBEaR: Computational Semantics Beyond Events and Roles

Organizers: *Eduardo Blanco and Roser Morante*

Venue: Bolden 4

Tuesday, June 5, 2018

08:00–17:00 **NAACL Registration**

09:00–09:10 **Opening Remarks**

09:10–09:30 Using Hedge Detection to Improve Committed Belief Tagging
Morgan Ulinski, Seth Benjamin, and Julia Hirschberg

09:30–10:00 Paths for uncertainty: Exploring the intricacies of uncertainty identification for news
Chrysoula Zerva and Sophia Ananiadou

10:00–10:20 Detecting Sarcasm is Extremely Easy ;-)
Natalie Parde and Rodney Nielsen

10:20–10:30 **Discussion Session 1**

10:30–11:00 **Coffee Break**

11:00–11:30 GKR: the Graphical Knowledge Representation for semantic parsing
Aikaterini-Lida Kalouli and Richard Crouch

11:30–12:20 Computational Argumentation: A Journey Beyond Semantics, Logic, Opinions, and Easy Tasks
Ivan Habernal

12:20–12:30 **Discussion Session 2**

Style-Var: 2nd Workshop on Stylistic Variation

Organizers: Julian Brooke, Lucie Flekova, Moshe Koppel, and Thamar Solorio

Venue: Bolden 3

Tuesday, June 5, 2018

08:00–17:00 NAACL Registration

09:15–09:30 Opening Remarks

09:30–10:30 **Invited Talk by James W. Pennebaker: Measuring Linguistic Variation with Function Words**

10:30–11:00 **Break**

11:00–11:30 Stylistic variation over 200 years of court proceedings according to gender and social class

Stefania Degaetano-Ortlib

11:30–12:00 Stylistic Variation in Social Media Part-of-Speech Tagging

Murali Raghu Babu Balusu, Taha Merghani, and Jacob Eisenstein

12:00–12:30 Detecting Syntactic Features of Translated Chinese

Hai Hu, Wen Li, and Sandra Kübler

12:30–14:00 **Lunch**

14:00–15:00 **Invited Talk by Rada Mihalcea: What Does Language Tell us about the People Behind it**

15:00–15:30 Evaluating Creative Language Generation: The Case of Rap Lyric Ghostwriting

Peter Potash, Alexey Romanov, and Anna Rumshisky

15:30–16:00 **Break**

16:00–17:00 **Invited Talk by Barbara Plank: Author Profiling from Text and Beyond**

17:00–17:30 Cross-corpus Native Language Identification via Statistical Embedding

Francisco Rangel, Paolo Rosso, Julian Brooke, and

Alexandra Uitdenbogerd

Story-NLP: Workshop on Storytelling

Organizers: *Margaret Mitchell, Kenneth Huang, Francis Ferraro, and Ishan Misra*

Venue: Bolden 6

Tuesday, June 5, 2018

08:00–17:00 **NAACL Registration**

10:00–10:10 **Opening Remarks**

10:15–11:00 **Invited Talk by Nasrin Mostafazadeh: Eventful Context Modeling: The Case of Story Comprehension and Story Generation**

11:00–11:25 Learning to Listen: Critically Considering the Role of AI in Human Storytelling and Character Creation
Anna Kasunic and Geoff Kaufman

11:25–11:50 Linguistic Features of Helpfulness in Automated Support for Creative Writing
Melissa Roemmele and Andrew Gordon

11:50–13:45 **Lunch**

13:45–14:10 A Pipeline for Creative Visual Storytelling
Stephanie Lukin, Reginald Hobbs, and Clare Voss

14:10–14:35 Telling Stories with Soundtracks: An Empirical Analysis of Music in Film
Jon Gillick and David Bamman

14:35–15:00 Towards Controllable Story Generation
Nanyun Peng, Marjan Ghazvininejad, Jonathan May, and Kevin Knight

15:00–15:20 **Break**

15:20–16:00 **Storytelling Challenge**

16:00–16:25 An Encoder-decoder Approach to Predicting Causal Relations in Stories
Melissa Roemmele and Andrew Gordon

16:25–16:50 Neural Event Extraction from Movies Description
Alex Tozzo, Dejan Jovanovic, and Mohamed R. Amer

16:50–17:00 **Closing Remarks**

SCLeM: Subword and Character LEvel Models in NLP

Organizers: *Manaal Faruqui, Hinrich Schütze, Isabel Trancoso, Yulia Tsvetkov, and Yadollah Yaghoobzadeh*

Venue: Bolden 6

Wednesday, June 6, 2018

08:00–16:00 NAACL Registration

09:30–09:45 Opening Remarks (Manaal Faruqui)

09:45–10:30 Invited Talk: Orthographic Social Variation in Online Writing (Jacob Eisenstein)

10:30–11:00 Coffee Break

11:00–11:45 Invited Talk: Not All That Glitters is Gold (Barbara Plank)

11:45–12:00 Best paper (sponsor: Microsoft Research) Talk #1

12:00–14:00 Lunch Break

14:00–14:45 Invited Talk: Morphology – When is it Useful in Neural Models? (Graham Neubig)

14:45–15:45 Poster Session & Coffee Break

- Morphological Word Embeddings for Arabic Neural Machine Translation in Low-Resource Settings
Pamela Shapiro and Kevin Duh
 - Entropy-Based Subword Mining with an Application to Word Embeddings
Ahmed El-Kishky, Frank Xu, Aston Zhang, Stephen Macke, and Jiawei Han
 - A Comparison of Character Neural Language Model and Bootstrapping for Language Identification in Multilingual Noisy Texts
Wafaa Adouane, Simon Dobnik, Jean-Philippe Bernardy, and Nasredine Semmar
 - Addressing Low-Resource Scenarios with Character-aware Embeddings
Sean Papay, Sebastian Padó, and Ngoc Thang Vu
 - Subword-level Composition Functions for Learning Word Embeddings
Bofang Li, Aleksandr Drozd, Tao Liu, and Xiaoyong Du
 - Discovering Phonesthemes with Sparse Regularization
Nelson F. Liu, Gina-Anne Levow, and Noah A. Smith
 - Meaningless yet meaningful: Morphology grounded subword-level NMT
Tamali Banerjee and Pushpak Bhattacharyya
 - Fast Query Expansion on an Accounting Corpus using Sub-Word Embeddings
Hrishikesh Ganu and Viswa Datha P
 - Incorporating Subword Information into Matrix Factorization Word Embeddings
Alexandre Salle and Aline Villavicencio
 - A Multi-Context Character Prediction Model for a Brain-Computer Interface
Shiran Dudy, Shaobin Xu, Steven Bedrick, and David Smith
- 15:45–16:30 Invited Talk: Romanization, Non-standard Orthography and Text Entry (Brian Roark)
- 16:30–16:45 Best paper (sponsor: Microsoft Research) Talk #2
- 16:45–17:30 Invited Talk: What Makes a Character-level Neural Model work? (Wang Ling)

TextGraphs: Twelfth Workshop on Graph-Based Methods for Natural Language Processing

Organizers: *Goran Glavaš, Swapna Somasundaran, Martin Riedl, and Eduard Hovy*

Venue: Bolden 1

Wednesday, June 6, 2018

08:00–16:00 **NAACL Registration**

09:00–09:10 **Opening Remarks**

Session 1

09:10–10:10 **Invited talk**

10:10–10:30 Scientific Discovery as Link Prediction in Influence and Citation Graphs
Fan Luo, Marco A. Valenzuela-Escárcega, Gus Hahn-Powell, and Mihai Surdeanu

10:30–11:00 **Coffee break**

Session 2

11:00–11:20 Efficient Generation and Processing of Word Co-occurrence Networks Using corpus2graph
Zheng Zhang, Pierre Zweigenbaum, and Ruqing Yin

11:20–11:40 Multi-hop Inference for Sentence-level TextGraphs: How Challenging is Meaningfully Combining Information for Science Question Answering?
Peter Jansen

11:40–12:05 Multi-Sentence Compression with Word Vertex-Labeled Graphs and Integer Linear Programming
Elvys Linhares Pontes, Stéphane Huet, Thiago Gouveia da Silva, Andréa Carneiro Linhares, and Juan-Manuel Torres-Moreno

12:05–14:05 **Lunch break**

Session 3

14:05–15:05 **Invited talk: Hierarchical Representation Learning on Graphs**

15:05–15:30 Large-scale spectral clustering using diffusion coordinates on landmark-based bipartite graphs
Khiem Pham and Guangliang Chen

15:30–16:00 **Coffee break**

Session 4

16:00–16:25 Efficient Graph-based Word Sense Induction by Distributional Inclusion Vector Embeddings
Haw-Shiuan Chang, Amol Agrawal, Ananya Ganesh, Anirudha Desai, Vinayak Mathur, Alfred Hough, and Andrew McCallum

16:25–16:50 Fusing Document, Collection and Label Graph-based Representations with Word Embeddings for Text Classification
Konstantinos Skianis, Fragkiskos Malliaros, and Michalis Vazirgiannis

16:50–17:05 Embedding Text in Hyperbolic Spaces
Bhuwan Dhingra, Christopher Shallue, Mohammad Norouzi, Andrew Dai, and George Dahl

17:05–17:15 **Closing remarks**

FigLang: Workshop on Figurative Language Processing

Organizers: *Beata Beigman Klebanov, Ekaterina Shutova, Patricia Lichtenstein, Smaranda Muresan, and Chee Wee*

Venue: Bolden 2

Wednesday, June 6, 2018

08:00–16:00 NAACL Registration

09:00–09:10 Opening remarks

09:10–10:10 Invited Talk: Tony Veale “When You Come To A Fork In The Road, Take It: Complementary Approaches to Metaphor Generation”

10:10–10:30 Challenges in Finding Metaphorical Connections
Katy Gero and Lydia Chilton

10:30–11:00 Coffee break

11:00–11:20 Linguistic Features of Sarcasm and Metaphor Production Quality
Stephen Skalicky and Scott Crossley

11:20–11:40 Leveraging Syntactic Constructions for Metaphor Identification
Kevin Stowe and Martha Palmer

11:40–12:00 Literal, Metaphorical or Both? Detecting Metaphoricity in Isolated Adjective-Noun Phrases
Agnieszka Mykowiecka, Małgorzata Marciniak, and Aleksander Wawer

12:00–12:20 Catching Idiomatic Expressions in EFL Essays
Michael Flor and Beata Beigman Klebanov

12:20–14:00 Lunch

14:00–14:20 Predicting Human Metaphor Paraphrase Judgments with Deep Neural Networks
Yuri Bizzoni and Shalom Lappin

14:20–14:40 A Report on the 2018 VUA Metaphor Detection Shared Task
Chee Wee (Ben) Leong, Beata Beigman Klebanov, and Ekaterina Shutova

14:40–15:40 Poster Session

- An LSTM-CRF Based Approach to Token-Level Metaphor Detection
Malay Pramanick, Ashim Gupta, and Pabitra Mitra
- Unsupervised Detection of Metaphorical Adjective-Noun Pairs
Malay Pramanick and Pabitra Mitra
- Phrase-Level Metaphor Identification Using Distributed Representations of Word Meaning
Omnia Zayed, John Philip McCrae, and Paul Buitelaar
- Bigrams and BiLSTMs Two Neural Networks for Sequential Metaphor Detection
Yuri Bizzoni and Mehdi Ghanimifard

- Computationally Constructed Concepts: A Machine Learning Approach to Metaphor Interpretation Using Usage-Based Construction Grammatical Cues
Zachary Rosen
- Neural Metaphor Detecting with CNN-LSTM Model
Chuhan Wu, Fangzhao Wu, Yubo Chen, Sixing Wu, Zhigang Yuan, and Yongfeng Huang
- Di-LSTM Contrast : A Deep Neural Network for Metaphor Detection
Krishnakant Swarnkar and Anil Kumar Singh
- Conditional Random Fields for Metaphor Detection
Anna Mosolova, Ivan Bondarenko, and Vadim Fomin
- Detecting Figurative Word Occurrences Using Recurrent Neural Networks
Agnieszka Mykowiecka, Aleksander Wawer, and Małgorzata Marciniak
- Multi-Module Recurrent Neural Networks with Transfer Learning
Filip Skurniak, Maria Janicka, and Aleksander Wawer
- Using Language Learner Data for Metaphor Detection
Egon Stemle and Alexander Onysko

15:40–16:00 **Coffee break**

16:00–17:00 **Invited Talk: Marilyn Walker “Hyperbole, Rhetorical Questions and Sarcasm: Figurative Language in Social Media”**

PEOPLES: Second Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media

Organizers: *Malvina Nissim, Viviana Patti, Barbara Plank, and Claudia Wagner*

Venue: Strand 11

Wednesday, June 6, 2018

08:00–16:00 NAACL Registration

08:50–09:00 Opening Remarks

09:00–10:30 Session 1

09:00–09:20 What makes us laugh? Investigations into Automatic Humor Classification
Vikram Ahuja, Taradheesh Bali, and Navjyoti Singh

09:20–09:40 Social and Emotional Correlates of Capitalization on Twitter
Sophia Chan and Alona Fyshe

09:40–10:00 Building an annotated dataset of app store reviews with Appraisal features in English and Spanish
Natalia Mora and Julia Lavid-López

10:00–10:15 Enabling Deep Learning of Emotion With First-Person Seed Expressions
Hassan Alhuzali, Muhammad Abdul-Mageed, and Lyle Ungar

10:15–10:30 A Dataset of Hindi-English Code-Mixed Social Media Text for Hate Speech Detection
Aditya Bohra, Deepanshu Vijay, Vinay Singh, Syed Sarfaraz Akhtar, and Manish Shrivastava

10:30–11:00 Coffee break

11:00–12:30 Session 2

11:00–12:00 The Social and the Neural Network: How to Make Natural Language Processing about People again
Dirk Hovy

12:00–12:15 Observational Comparison of Geo-tagged and Randomly-drawn Tweets
Tom Lippincott and Annabelle Carrell

12:15–12:30 Johns Hopkins or johnny-hopkins: Classifying Individuals versus Organizations on Twitter
Zach Wood-Doughty, Praateek Mahajan, and Mark Dredze

12:30–14:00 Lunch

14:00–15:30 Session 3

14:00–15:00 The Potential of the Computational Linguistic Analysis of Social Media for Population Studies
Letizia Mencarini

15:00–15:15 Understanding the Effect of Gender and Stance in Opinion Expression in Debates on “Abortion”
Esin Durmus and Claire Cardie

15:15–15:30 Frustrated, Polite, or Formal: Quantifying Feelings and Tone in Email
Niyati Chhaya, Kushal Chawla, Tanya Goyal, Projjal Chanda, and Jaya Singh

15:30–16:00 **Coffee break**

16:00–17:30 **Session 4**

16:00–16:20 Reddit: A Gold Mine for Personality Prediction
Matej Gjurković and Jan Šnajder

16:20–16:40 Predicting Authorship and Author Traits from Keystroke Dynamics
Barbara Plank

16:40–17:00 Predicting Twitter User Demographics from Names Alone
Zach Wood-Doughty, Nicholas Andrews, Rebecca Marvin, and Mark Dredze

17:00–17:15 Modeling Personality Traits of Filipino Twitter Users
Edward Tighe and Charibeth Cheng

17:15–17:30 Grounding the Semantics of Part-of-Day Nouns Worldwide using Twitter
David Vilares and Carlos Gómez-Rodríguez

17:30–18:00 **Discussion and Closing**

CRAC: Workshop on Computational Models of Reference, Anaphora and Coreference

Organizers: *Massimo Poesio, Vincent Ng, and Maciej Ogródniczuk*

Venue: Bolden 3

Wednesday, June 6, 2018

08:00–16:00 NAACL Registration

09:00–10:30 Session 1

09:00–09:10 Welcome (Massimo Poesio, Vincent Ng, Maciej Ogródniczuk)

09:10–10:00 Invited Talk (Ana Marasovic)

10:00–10:30 Anaphora Resolution for Twitter Conversations: An Exploratory Study
Berfin Aktaş, Tatjana Scheffler, and Manfred Stede

10:30–11:00 Break

11:00–12:30 Session 2: Shared Task, Plural Reference

11:00–11:30 Anaphora Resolution with the ARRAU Corpus

Massimo Poesio, Yulia Grishina, Varada Kolhatkar, Nafise Moosavi, Ina Roesiger, Adam Roussel, Fabian Simonjetz, Alexandra Uma, Olga Uryupina, Juntao Yu, and Heike Zinsmeister

11:30–12:00 Rule- and Learning-based Methods for Bridging Resolution in the ARRAU Corpus

Ina Roesiger

12:00–12:30 A Predictive Model for Notional Anaphora in English
Amir Zeldes

12:30–14:00 Lunch

14:00–15:30 Session 3: Bridging, Discourse deixis, Anaphora in German, Corpus annotation 1

14:00–14:20 Integrating Predictions from Neural-Network Relation Classifiers into Coreference and Bridging Resolution
Ina Roesiger, Maximilian Köper, Kim Anh Nguyen, and Sabine Schulte im Walde

14:20–14:50 Towards Bridging Resolution in German: Data Analysis and Rule-based Experiments
Janis Pagel and Ina Roesiger

14:50–15:10 Detecting and Resolving Shell Nouns in German
Adam Roussel

15:10–15:30 PAWS: A Multi-lingual Parallel Treebank with Anaphoric Relations
Anna Nedoluzhko, Michal Novák, and Maciej Ogródniczuk

15:30–16:00 Break

16:00–17:30 Session 4: Corpus Annotation 2, Cognitive Models

16:00–16:30 A Fine-grained Large-scale Analysis of Coreference Projection
Michał Novák

16:30–17:00 Modeling Brain Activity Associated with Pronoun Resolution in English and Chinese
Jixing Li, Murielle Fabre, Wen-Ming Luh, and John Hale

17:00–17:30 Event versus entity co-reference: Effects of context and form of referring expression
Sharid Loáiciga, Luca Bevacqua, Hannah Rohde, and Christian Hardmeier

SpLU: First International Workshop on Spatial Language Understanding

Organizers: *Parisa Kordjamshidi, Archna Bhatia, James Pustejovsky, and Marie-Francine Moens*

Venue: Bolden 4

Wednesday, June 6, 2018

08:00–16:00 NAACL Registration

Session 1

09:00–09:10 Opening remarks (**Parisa Kordjamshidi**)

09:10–10:10 Keynote talk: Natural Language Acquisition and Grounding for Embodied Robotic Systems (**Anthony G. Cohn**)

10:10–10:30 Exploring the Functional and Geometric Bias of Spatial Relations Using Neural Language Models
Simon Dobnik, Mehdi Ghanimifard, and John Kelleher

10:30–11:00 Coffee Break

Session 2

11:00–11:20 Building and Learning Structures in a Situated Blocks World Through Deep Language Understanding
Ian Perera, James Allen, Choh Man Teng, and Lucian Galescu

11:20–11:40 Computational Models for Spatial Prepositions
Georgiy Platonov and Lenhart Schubert

11:40–12:00 Lexical Conceptual Structure of Literal and Metaphorical Spatial Language: A Case Study of “Push”
Bonnie Dorr and Mari Olsen

12:00–12:20 Representing Spatial Relations in FrameNet
Miriam R L Petruck and Michael J Ellsworth

12:20–14:10 Lunch Break

Session 3

14:10–15:10 Keynote talk: Understanding Spatial Expressions (**James F. Allen**)

15:10–15:30 Points, Paths, and Playscapes: Large-scale Spatial Language Understanding Tasks Set in the Real World
Jason Baldridge, Tania Bedrax-Weiss, Daphne Luong, Srini Narayanan, Bo Pang, Fernando Pereira, Radu Soricut, Michael Tseng, and Yuan Zhang

15:30–16:00 Coffee Break

Session 4

16:00–16:20 Anaphora Resolution for Improving Spatial Relation Extraction from Text
Umar Manzoor and Parisa Kordjamshidi

16:20–16:40 The Case for Systematically Derived Spatial Language Usage
Bonnie Dorr and Clare Voss

16:40–17:30 Panel (**James Pustejovsky, Marie-Francine Moens, James F. Allen, Bonnie Dorr, Anthony G. Cohn**)

17:30–17:40 Concluding remarks (**Archna Bhatia**)

Anti-Harassment Policy

The open exchange of ideas, the freedom of thought and expression, and respectful scientific debate are central to the aims and goals of a ACL conference. These require a community and an environment that recognizes the inherent worth of every person and group, that fosters dignity, understanding, and mutual respect, and that embraces diversity. For these reasons, ACL is dedicated to providing a harassment-free experience for participants at our events and in our programs.

Harassment and hostile behavior are unwelcome at any ACL conference. This includes: speech or behavior (including in public presentations and on-line discourse) that intimidates, creates discomfort, or interferes with a person's participation or opportunity for participation in the conference. We aim for ACL conferences to be an environment where harassment in any form does not happen, including but not limited to: harassment based on race, gender, religion, age, color, national origin, ancestry, disability, sexual orientation, or gender identity. Harassment includes degrading verbal comments, deliberate intimidation, stalking, harassing photography or recording, inappropriate physical contact, and unwelcome sexual attention.

It is the responsibility of the community as a whole to promote an inclusive and positive environment for our scholarly activities. In addition, any participant who experiences harassment or hostile behavior may contact any current member of the ACL Board or contact Priscilla Rasmussen, who is usually available at the registration desk of the conference. Please be assured that if you approach us, your concerns will be kept in strict confidence, and we will consult with you on any actions taken.

The ACL board members are listed at:

<https://www.aclweb.org/website/about>

The full policy and its implementation is defined at:

https://www.aclweb.org/adminwiki/index.php?title=Anti-Harassment_Policy

8

Local Information

Welcome to New Orleans! The NAACL HLT 2018 conference will be held at the Hyatt Regency New Orleans Hotel and this will also be the main conference hotel.

Hyatt Regency New Orleans

601 Loyola Avenue

New Orleans, Louisiana, USA, 70113

Tel: +1 504 561 1234

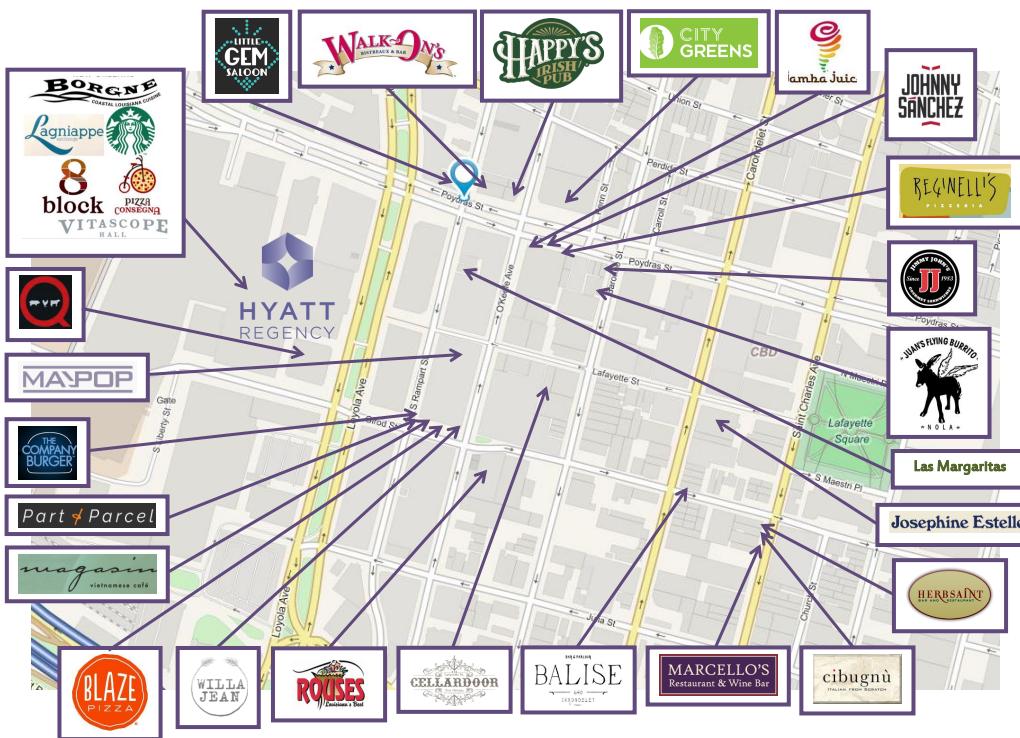
<https://neworleans.regency.hyatt.com/en/hotel/home.html>

The hotel is located adjacent to the Superdome and just minutes from many fun things to do in New Orleans. Stroll, meander, wander, or take the Streetcar to the nearby French Quarter and explore exciting events. In downtown New Orleans, the hotel is conveniently located near many popular attractions which are accessible either by walking or the city streetcars. Valet parking is offered on property and the hotel is attached to a public self-parking garage.

Local Guide Be sure to explore this fascinating city in your spare time! Local attractions can be found at the conference website at http://naacl2018.org/local_info.html.

Local Public Transportation Guests can enjoy quick and easy transportation to the city's most popular attractions via the Loyola-UPT Streetcar Line, which passes approximately every 20 minutes across from the hotel. The cost to ride streetcars in New Orleans is \$1.25 and can be paid with exact change when you board. 1-Day and 3-Day unlimited ride Jazzy Passes are also available for \$3 and \$9. For more information, please visit the hotel's Concierge.

Safety While there are no areas within close walk of the hotel to caution against, as with any big city, guests should walk in groups, especially if they are walking around late at night, and ensure they are paying attention to their belongings.



<i>Casual</i>			
<i>American</i>			
Walk – On's	1009 Poydras st.	309.6530	Boast great steaks, killer seafood and a variety of South Louisiana specialties
Port Of 'Call	838 Esplanade	523.0120	Famous for burgers and steaks.
Snug Harbor	626 Frenchman St.	949.0696	Live Jazz and great regional cooking.
Crescent City Brewhouse	527 Decatur	522.0571	The only French Quarter micro-brewery
<i>Barbecue</i>			
Q Smokery & Cafe	639 Loyola Ave.	525.4044	World Champion BBQ, Burgers, Salads, and more.
The Joint	701 Mazant St.	949.3232	World famous slow cooked pulled pork and beef brisket
Voodoo Barbecue	1501 St. Charles St.	522.4647	Dry rubbed meats with amazing local spices.
<i>Cajun</i>			
Bon Ton	401 Magazine St.	524.3386	Authentic Cajun Cuisine
Jacques-Imo's	8324 Oak St.	861.0886	Eclectic mix of Creole and Cajun specialties
Mulaté's	201 Julia St.	522.1491	Live Cajun music and dancing with authentic Cajun food.
<i>Creole Contemporary</i>			
Palace Café	605 Canal St.	482.9179	"Flavor of New Orleans" cuisine in an upbeat and lively grand café.
<i>Dessert & Coffee</i>			
Café Du Monde	800 Decatur	525.4544	Famous for Beignets
<i>Indian</i>			
Nirvana	4308 Magazine	894.9797	Inspired by curry houses in Europe.
<i>Italian</i>			
Adolfo's	611 Frenchmen St.	948.3800	Nice variations on their meats, pasta, and sauces.
Louisiana Pizza Kitchen	95 French Market Pl.	522.0571	Dishes influenced by European and Mediterranean tradition.
Pascales Manale	1838 Napoleon Ave.	895.4877	Famous for creating barbecued shrimp.
<i>Seafood</i>			
Acme Oyster House	724 Iberville	522.5973	Fresh, hand-shucked Louisiana Oysters.
Felix Oyster	739 Iberville St.	522.4440	Delicious Creole and Cajun cuisine
Deanie's	841 Iberville	581.1316	Authentic and unique New Orleans culinary traditions.
Red Fish Grill	115 Bourbon St.	598.1200	A triumph of cuisine, style, and design

9

On-Site Childcare

NAACL 2018 will be offering on-site childcare at the conference hotel by advanced reservation or walk-in. Walk-ins are subject to availability, but there should be plenty of space, so please bring your kids! The cost of the childcare is partially subsidized by the NAACL. The cost for general registrants will be \$80 USD per day per child. For student registrants the cost will be \$40 USD per day per child.

The childcare rooms are located in Strand 3 and 4 which are on the 2nd Level (Strand Foyer), across from the stairs and escalators. The care is available for children 6 months to 12 years old from 8:30 a.m. to 6 p.m. during the main conference, plus the tutorials and workshop days. The caregivers are from a professional childcare company, KiddieCorp, all of whom are CPR/first aid certified and with child care experience. The childcare has activities and play materials such as arts and crafts as well as appropriate toys for babies as well as a quiet area for napping.

In addition to childcare, we invite you to bring your children, spouses and loved ones to the social event at the conference. You can purchase additional meals or additional tickets to the social event for your family members. The social event for this year is an evening at the Mardi Gras World museum. It's a family friendly venue, and we hope to see you and your family there!

We are grateful to Amazon for its generous support of the childcare at NAACL 2018.

Below is information from KiddieCorp about their services.

Hello NAACL Parents!

Thank you very much for your interest in NAACL's children's program. Our goal is to provide your children with a program they want to attend, while providing you with that critical "peace of mind" feeling so you can attend the conference activities without worrying.

KiddieCorp is pleased to provide a children's program during NAACL 2018. KiddieCorp has more than thirty years of experience providing high quality children's programs and youth services to conventions, trade shows and special events. We take caring for your children very seriously. KiddieCorp has enjoyed a long-time partnership with the American Academy of Pediatrics, which has helped to establish KiddieCorp as a premier provider of event children's program services. This is the second year we have provided services for ACL conferences—we started last year at ACL 2017 in Vancouver.

Activities Activities include exciting themes, arts and crafts, group games, music and movement, board games, story time, dramatic play, etc. We provide activities appropriate for each age group, using safe, sturdy equipment that you can feel comfortable with. Children can make their own choices within KiddieCorp's program.

Commitment Our goal is to provide your children with a comfortable, safe and happy experience. Our staff to child ratios are high to ensure that every child feels special (1:2 for children ages 6 months through 11 months old; 1:3 for children ages 1 through 2 years old; 1:5 for children ages 3 through 5 years old; 1:7 for children ages 6 through 12 years old). KiddieCorp team members are selected according to their integrity, experience, education and enthusiasm. They must be wonderful with kids! In addition to our competitive and selective hiring process, KiddieCorp remains at the top of the industry by carrying ample liability insurance.

Where, When, For Whom The program is for children ages 6 months through 12 years old. The dates for the program are June 1 to June 6, 2018 and the program will be located in Strand 3 and 4 at the Hyatt Regency New Orleans hotel. Snacks and beverages will be provided and meals need to be supplied by parents when checking in your child each day.

You will also have the option to purchase lunch for your child(ren) at registration. The cost will be around \$15. Lunch must be ordered by 10:00am each day and paid in cash only.

Other Info

- Please label your child's belongings. We will maintain a lost and found, however, KiddieCorp does not accept responsibility for the loss or theft of any toy, book, or other personal items.
- For parents with infants, please bring diaper changing supplies, formula/baby food, and a change of clothes.
- Cancellation Policy: Cancellations must be made to KiddieCorp prior to May 1, 2018 for a full refund. Cancellations made after that date will be subject to a 50% cancellation fee. Once the program has begun, no refunds will be issued.
- KiddieCorp staff does not administer medication. To ensure a safe and fun-filled environment, any child who is ill will not be admitted to the children's program.

Need more information? KiddieCorp is always available to answer any questions. Feel free to contact KiddieCorp's on-site manager, Priscilla Maxwell. Her cell phone number is 1-828-545-9342.

Example activities - Space Explorer Children will participate in different group games, arts & crafts, reading, and much more! Five, Four, Three, Two, One, Blast off! It's time to go to KiddieCorp Space Explorer Camp. Bounce into our space dome tents and let our creativity take off.

Let's imagine what zero gravity feels like when we make our own pair of moon boots out of paper bags. The masterpieces we create at the art table can include play dough alien life forms, foam planet bracelets and paper plate spaceships.

When we get back on earth, we can rediscover gravity by tossing the balls in the air with the parachute. We'll also have a chance to play astronaut ring toss, musical planets, and moon, moon, star (our version of duck, duck, goose). So let's buckle our seat belts, prepare for countdown, and blast off to KiddieCorp Space Explorer Camp.

Index

- Abate, Solomon Teferra, 17
Abdou, Mostafa, 154, 162
Abdu, Haruna, 17
Abdul-Mageed, Muhammad, 183
Abdullah, Malak, 155
Abend, Omri, 11, 48, 76
Abreu, Carla, 158
AbuRa'ed, Ahmed, 170
Acharya, Sabita, 70
Adouane, Wafia, 179
Aedmaa, Eleri, 61
Afrin, Tazin, 171
Aggarwal, Srishti, 18
Agirrezabal, Manex, 11
Agrawal, Amol, 180
Aguilar, Gustavo, 107
Ahmad, Tariq, 154
Ahmed, Eisha, 135
Ahmed, Hanady, 154
Ahmed, Usman, 158
Ahuja, Vikram, 183
Aina, Laura, 152
Aizawa, Akiko, 160
Akhtar, Md Shad, 64
Akhtar, Syed Sarfaraz, 128, 183
Akkasi, Abbas, 160
Akkaya, Cem, 11, 12
Aktaş, Berfin, 185
Al Hanai, Tuka, 144
Albiero, Beatriz, 18
Aldarmaki, Hanan, 62
Alexa, Larisa, 156
Alfano, Mark, 173
Alfonseca, Enrique, 11
Alfter, David, 169, 171
Alhuzali, Hassan, 183
Alikaniotis, Dimitrios, 11
Allamanis, Miltiadis, 15, 21
Allen, James, 167, 186
Alm, Cecilia Ovesdotter, 140
Altsyler, Edgar, 174
Amarandei, Sandra Maria, 153
Ambroselli, Carl, 122
Amer, Mohamed R., 147, 178
Amiri, Hadi, 139
Ammar, Waleed, 12, 47, 49, 128
Amoia, Marilisa, 36
Amorim, Evelin, 46
Amplayo, Reinald Kim, 77
An, Bo, 84
Ananiadou, Sophia, 116, 176
Anastasopoulos, Antonios, 35
Andersen, Øistein E., 171
Anderson, Tim, 11
Andrassy, Bernt, 95
Andreas, Jacob, 136, 145
Andrews, Nicholas, 184
Angel, Jason, 171
Angeli, Gabor, 12
Angelidis, Stefanos, 55
Angelova, Galia, 157
Antoine, Jean-Yves, 91
Antoniak, Maria, 80
Aono, Masaki, 153
Apidianaki, Marianna, 39, 45, 105, 152
Aragaw, Bitsseat, 17
Arampatzis, Avi, 156
Ardati, Amer, 70

- Arora, Richa, 46
Aroyehun, Segun Taofeek, 171
Arroyo-Fernández, Ignacio, 162
Arseniev-Koehler, Alina, 174
Arshad Islam, Muhammad, 158
Artstein, Ron, 18
Artzi, Yoav, 77, 116, 150
Asamoah Owusu, Dennis, 156
Assylbekov, Zhenisbek, 106
Atanasov, Atanas, 157
Attia, Mohammed, 162
Au, Rhoda, 144
Aufrant, Lauriane, 89
Augenstein, Isabelle, 89, 135, 155
Auli, Michael, 11, 40, 48
Axelrod, Amitai, 16
Axtmann, Nico, 49, 71
Ayalew, Tensae, 17
Azab, Mahmoud, 146
Aziz, Wilker, 11, 80
- Babaoglu, Ismail, 153
Badaro, Gilbert, 154, 166
Bahdanau, Dzmitry, 168
Baiju, Aiswarya, 19
Bakarov, Amir, 166
Baker, Simon, 38
Balasubramanian, Niranjan, 168
Balazs, Jorge, 157
Baldridge, Jason, 186
Baldwin, Timothy, 11, 64, 101, 137, 164
Bali, Taradheesh, 183
Ballesteros, Miguel, 71, 136, 152
Balusu, Murali Raghu Babu, 177
Baly, Ramy, 42, 84
Bamman, David, 36, 178
Banerjee, Tamali, 179
Bangalore, Srinivas, 6
Bansal, Mohit, 11, 17, 20, 35, 78, 81, 125, 131, 139, 145
Bao, Forrest Sheng, 127
Barak, Libby, 16, 18
Barbieri, Francesco, 136, 152, 166
Barbosa, Denilson, 34
Barik, Biswanath, 161
Barnwal, Dhiraj, 50
Baron, Alistair, 158
Baroni, Marco, 98
Barrett, Maria, 140, 169
Barriere, Caroline, 159
Barrón-Cedeño, Alberto, 65
Bartley, G. Keith, 175
Barzilay, Regina, 54
Basile, Angelo, 156
Basile, Valerio, 11, 152, 167
Basilici, Roberto, 11
- Bastings, Joost, 105
Basu Roy Chowdhury, Somnath, 38
Bateman, Leila, 175
Bawden, Rachel, 102
Baziotis, Christos, 154, 156, 158
Beauchamp, Nicholas, 55
Beaulieu, Jonathan, 156
Béchet, Nicolas, 161
Bechikh Ali, Chedi, 153
Bedrax-Weiss, Tania, 186
Bedrick, Steven, 18, 179
Beigman Klebanov, Beata, 11, 46, 144, 169, 181
Bekki, Daisuke, 84, 89
Belding, Elizabeth, 47
Belfer, Robert, 135
Belinkov, Yonatan, 105
Bellot, Patrice, 157
Beltagy, Iz, 49
Bengio, Yoshua, 168
Benjamin, Seth, 176
Benton, Adrian, 54
Berant, Jonathan, 11, 68, 77, 100, 124, 165
Berend, Gábor, 161
Berenstein, Ariel J., 174
Berg-Kirkpatrick, Taylor, 146
Bermanis, Toms, 106
Berio, Giuseppe, 161
Bernardi, Raffaella, 56
Bernardy, Jean-Philippe, 179
Bernhard, Delphine, 11
Bernier-Colborne, Gabriel, 159
Berry, Danielle, 82
Berzak, Yevgeni, 139
Besacier, Laurent, 16
Beschke, Sebastian, 165
Beshah, Tibebe, 17
Bestgen, Yves, 171
Bethard, Steven, 152, 166
Betke, Margrit, 18
Bevacqua, Luca, 185
Bhagat, Rahul, 11
Bhagavatula, Chandra, 47, 49
Bhanderi, Shruti, 135
Bhat, Irshad, 91
Bhat, Riyaz A., 91
Bhat, Suma, 89
Bhatia, Archna, 186
Bhattacharyya, Pushpak, 63, 64, 76, 170, 179
Bhogadhi, Saroja, 65
Biemann, Chris, 128, 162, 169
Bikel, Dan, 34
Bin, Mao, 160
Bingel, Joachim, 169, 170
Birch, Alexandra, 102
Bisk, Yonatan, 168

- Bizzoni, Yuri, 181
Bjerva, Johannes, 89, 155, 170
Blanco, Eduardo, 58, 60, 176
Blodgett, Su Lin, 90
Bloom, David, 163
Blunsom, Phil, 90
Boghrati, Reihane, 12
Bohnet, Bernd, 11
Bohra, Aditya, 128, 183
Bojanowski, Piotr, 98
Boleda, Gemma, 152
Bollegalá, Danushka, 42, 59, 165, 167
Boltuzic, Filip, 164
Bondarenko, Ivan, 182
Bonial, Claire, 18
Bonin, Francesca, 133, 147
Bosselut, Antoine, 44, 129
Botschen, Teresa, 116, 167
Bouamor, Houda, 19
Bouchekif, Abdesselam, 153
Boudin, Florian, 132
Bowden, Kevin, 43
Bowman, Samuel, 46, 96, 101, 126, 137, 168
Boyd, Andrew, 70
Boyd-Graber, Jordan, 95, 145
Boyer, Kristy, 171
Brahmana, Chathurika, 160
Brassard, Ana, 164
Braune, Fabienne, 59
Bravo-Marquez, Felipe, 152
Brehm, Oliver, 170
Brendoerfer, Michael, 49, 71
Brew, Chris, 11, 161
Briscoe, Ted, 47, 171
Brodkin, Edward S., 175
Brooke, Julian, 137, 177
Bruno, James, 169
Brust, Chris, 169
Bryant, Christopher, 171
Brychcín, Tomáš, 161
Buckman, Jacob, 101
Buddhitha, Prasadith, 175
Budzianowski, Paweł, 141
Buechel, Sven, 135
Buffone, Anneke, 175
Bui, Hung, 99
Bui, Trung, 99, 129
Buitelaar, Paul, 181
Buljan, Maja, 59
Burstein, Jill, 169
Buscaldi, Davide, 159
Butnaru, Andrei, 170
Butzke, Daniel, 160
Buys, Jan, 90
Byrne, Bill, 71, 104
Çetinoğlu, Özlem, 12
Çöltekin, Çağrı, 12, 152
C. de Souza, José G., 45
Cabrio, Elena, 167
Cahill, Aoife, 11, 48
Cai, Liwei, 116
Cakici, Ruket, 11
Calixto, Iacer, 12
Callison-Burch, Chris, 39, 45, 105
Calvo, Rafael A., 174
Camacho-Collados, Jose, 15, 25, 152, 159, 166
Camargo e Souza Câmara, Igor de, 18
Cambria, Erik, 63, 143
Cameron, Richard, 70
Campillos, Leonardo, 12
Can, Ethem F., 153
Cançado, Marcia, 46
Cao, Di, 143
Cao, Jinghua, 154
Cao, Kai, 12
Cao, Yuan, 11
Card, Dallas, 128
Cardie, Claire, 19, 80, 87, 99, 183
Carenini, Giusepe, 11
Carpuat, Marine, 17, 67, 83, 88, 117, 152, 162
Carrell, Annabelle, 183
Carvalho, Vitor, 86
Casanueva, Iñigo, 15, 27, 141
Catt, Michael, 157
Celikyilmaz, Asli, 11, 44, 129
Cercas Curry, Amanda, 16, 173
Chambers, Nathanael, 127
Chan, Sophia, 183
Chan, Yee Seng, 12
Chanda, Projal, 183
Chandrasekaran, Arjun, 145
Chang, Haw-Shiuan, 59, 180
Chang, Kai-Wei, 42, 80, 95
Chang, Shiyu, 100
Chang, Walter, 129
Chang, Yi, 57
Changhai, Tian, 160
Chao, Wei-Lun, 56
Chao, Wenhan, 121
Charnois, Thierry, 159
Chatterjee, Rajen, 78
Chaturvedi, Snigdha, 47, 132
Chauhan, Geeticka, 16, 160
Chawla, Kushal, 183
Che, Wanxiang, 91
Chen, Bo, 84
Chen, Cen, 127
Chen, Cheng, 11
Chen, Chia-Ping, 154
Chen, Francine, 120
Chen, Guangliang, 180

- Chen, Guanliang, 171
Chen, Guanyi, 162
Chen, Henry Y., 152
Chen, Hsin-Hsi, 162
Chen, Huadong, 104
Chen, Hubie, 108
Chen, Jiajun, 104
Chen, Jiangui, 155
Chen, Jing, 156
Chen, Jin, 171
Chen, Ken, 34
Chen, Liheng, 34
Chen, Mingda, 143
Chen, Moshai, 161
Chen, Peng-Yu, 47
Chen, Tsung-Chieh, 144
Chen, Wei, 102, 156
Chen, Wenhui, 120, 130
Chen, Wenqing, 142
Chen, Xie, 171
Chen, Xilun, 99
Chen, Yan-Ying, 120
Chen, Yining, 150
Chen, Yubo, 182
Chen, Yun-Nung, 44, 120, 144
Chen, Zhimin, 164
Cheng, Charibeth, 184
Cheng, Fei, 121
Cheng, Hao, 148
Cheng, Pengxiang, 86
Cheng, Weiwei, 167
Cheng, Xueqi, 85
Cheng, Yu, 100
Chersoni, Emmanuele, 162
Cheung, Jackie Chi Kit, 62, 70, 165, 166, 168
Chhaya, Niyati, 183
Chi, Zewen, 155
Chiang, David, 35, 39, 104
Chilton, Lydia, 181
Chinkina, Maria, 169
Chinnappa, Dhivya, 60
Chioma, Enemoh, 18, 69
Chiticariu, Laura, 49
Cho, Jaemin, 120
Cho, Kyunghyun, 126
Cho, Won Ik, 157
Chodorow, Martin, 67
Choi, HongSeok, 159
Choi, Jihun, 163, 166
Choi, Jinho D., 140, 152
Choi, Yejin, 44, 129, 148, 168
Chollampatt, Shamil, 11
Choshen, Leshem, 48
Choubey, Prafulla Kumar, 85
Chouikha, Mohamed, 17
Christodoulopoulos, Christos, 85
Chronopoulou, Alexandra, 154
Chu-Carroll, Jennifer, 6, 93
Chung, Tagyoung, 82
Chung, Yu-An, 125
Ciobanu, Alina Maria, 162
Cirik, Volkan, 146
Clark, Christopher, 150
Clark, Elizabeth, 148, 150
Clark, Peter, 126
Coates, Joshua, 59
Cocos, Anne, 39, 105
Cohan, Arman, 129, 160, 175
Cohen, Adir, 100
Cohen, Shay B., 57, 96, 132
Cohen, William, 42
Cohn, Anthony G., 186
Cohn, Trevor, 64, 101, 137
Cohn-Gordon, Reuben, 81
Cola, Meredith, 175
Cole, Jeremy, 140
Collins, Michael, 12, 39, 97, 110
Conneau, Alexis, 56
Conrad, Stefan, 163
Cook, Paul, 162
Copestake, Ann, 168
Coppersmith, Glen, 175
Cordeiro, Silvio Ricardo, 12
Corpas Pastor, Gloria, 162
Coster, Joël, 156
Cotterell, Ryan, 34, 107
Crane, Matt, 69
Crawford, Miles, 49
Crook, Paul, 11
Crossley, Scott, 181
Crouch, Richard, 166, 176
Cui, Xia, 167
Cui, Xiaodong, 11
Cummins, Ronan, 171

da San Martino, Giovanni, 11
Daelemans, Walter, 124
Dagan, Ido, 87, 96, 124
Dahl, George, 180
Dai, Andrew, 11, 180
Dai, Bo, 12
Dai, Xin-Yu, 104
Dai, Zeyu, 42
Dalen, Reinder Gerard van, 156
Dallmann, Alexander, 160
Dalvi, Bhavana, 126, 128
Dalvi, Fahim, 105
Daly, Elizabeth, 133
Damonte, Marco, 57, 96
Danforth, Douglas, 169
Danilevsky, Marina, 49
Darari, Fariz, 152

- Datla, Vivek, 116
Daumé III, Hal, 54, 168, 174
Daval-Ferot, Guillaume, 153
Dave, Neisarg, 171
Davis, Larry, 145
Davis, Rafe, 19
Davis, Tyler, 174
Daxenberger, Johannes, 42, 65
De Bruyne, Luna, 153
De Clercq, Orphee, 153
De Hertog, Dirk, 171
De Kuthy, Kordula, 170
de Melo, Gerard, 11
de Souza, Jose, 11
de Wit, Simon, 156
Dearden, Edward, 158
Degaetano-Ortlib, Stefania, 177
Dehghani, Nazanin, 95
Deleris, Lea, 133, 147
Deleu, Johannes, 175
Delli Bovi, Claudio, 159
Delmonte, Rodolfo, 11
DeLucia, Luke, 147
Demberg, Vera, 165
Demeester, Thomas, 175
Demner-Fushman, Dina, 11
Deng, Haohui, 88
Deng, Jia, 146
Deng, Li, 101
Deng, Lingjia, 11
dePaiva, Valeria, 18
Deparis, Stephane, 133
Derczynski, Leon, 163
Dernoncourt, Franck, 129, 160
Desai, Anirudha, 180
Desmet, Bart, 175
Desmet, Piet, 170
Develder, Chris, 175
Devlin, Jacob, 39
Dey, Lipika, 155
Dharmaretnam, Dhanush, 146
Dhingra, Bhuwan, 42, 125, 180
Dhyani, Dushyanta, 160
di Buono, Maria Pia, 160
Di Eugenio, Barbara, 19, 70
Diab, Mona, 62, 99, 102, 106
Dickens, Carolyn, 70
Dickinson, Markus, 170
Diehl, Frank, 36
Ding, Guanqi, 154
Ding, Haibo, 135
Ding, Nan, 117
Ding, Peng, 163, 164
Ding, Xiao, 63
Dingwall, Nicholas, 60
Dinh, Dien, 17
- Dinu, Liviu P., 162
Diligach, Dmitriy, 166
Dobnik, Simon, 179, 186
Dogga, Pradeep, 50
Dong, Zhenyuan, 154
Doran, Christine, 40
Dorr, Bonnie, 186
Doshi-Velez, Finale, 137
Downey, Doug, 49
Doyle, Gabriel, 43, 70
Dragoni, Mauro, 153, 157, 160
Dragut, Eduard, 65
Dras, Mark, 11, 92
Dredze, Mark, 54, 183, 184
Drozd, Aleksandr, 179
Drozdov, Andrew, 137
Du, Pan, 155
Du, Xiaoyong, 179
Duan, Junwen, 63
Duan, Nan, 124
Dudy, Shiran, 18, 179
Duh, Kevin, 100, 166, 179
Duke, Jon, 95
Dukkipati, Ambedkar, 38
Dumitru, Bogdan, 162
Dunkelberger, Jason, 49
Dunn Lopez, Karen, 70
Duppada, Venkatesh, 152
Dürlich, Luise, 158
Durmus, Esin, 80, 183
Durrani, Nadir, 105
Durrett, Greg, 77
Dutt, Ritam, 50
Dutta, Rina, 174
Dušek, Ondřej, 11, 45
Dyer, Chris, 107
Dymetman, Marc, 11
Dziri, Nouha, 43
- Eckmann, Stefanie, 58
Eder, Tobias, 59
Edessa, Adey, 17
Edunov, Sergey, 40
Edwards, Erik, 49, 71
Effrosynidis, Dimitrios, 156
Eger, Steffen, 11, 42, 65
Egg, Markus, 11
Eisenstein, Jacob, 55, 95, 177, 179
Eisner, Jason, 34, 90, 107
Ekbal, Asif, 63, 64, 76
El Asri, Layla, 11
El Jundi, Obeida, 154
El-Hajj, Wassim, 154, 166
El-Kishky, Ahmed, 179
Elango, Venkatesh, 155
Elgohary, Ahmed, 49

- Elkan, Charles, 40
Ellinas, Nikolaos, 154, 156, 158
Ellsworth, Michael J., 186
Elsafy, Ahmed, 128
Elsahar, Hady, 45, 131
Elsayed, Ahmed, 152
ElSherief, Mai, 47
Emami, Ali, 62, 70
Embar, Varsha, 168
Erdmann, Alexander, 90
Erk, Katrin, 58, 77, 86
Escalante, Hugo Jair, 99
Espinosa Anke, Luis, 15, 25, 88, 152, 159
Etzioni, Oren, 49
Evanini, Keelan, 11
Evans, Richard, 157
Ezeani, Ignatius, 18, 69
Ezen-Can, Aysu, 153

Fabre, Murielle, 185
Fairon, Cédrick, 170
Fan, Chuchu, 166
Fan, Zhihao, 154
Fang, Hao, 148
Faour, Ahmad, 161
Farag, Youmna, 47
Farra, Noura, 12
Farri, Oladimeji, 116
Faruqui, Manaal, 162, 179
Federico, Marcello, 17
Feldman, Sergey, 47, 49
Felix, Matthieu, 104
Feng, Yansong, 44
Ferguson, Emily, 175
Ferguson, James, 87
Fern, Xiaoli, 18, 116
Fernandez Slezak, Diego, 174
Fernández-González, Daniel, 88, 138
Ferraro, Francis, 161, 178
Ferret, Olivier, 11
Filippova, Katja, 11
Fine, Alex, 175
Finegan-Dollak, Catherine, 129
Finin, Tim, 161
Finlayson, Mark A., 16
Finley, Gregory, 49, 71
Fiorini, Nicolas, 122
Fleşcan-Lovin-Arseni, Iuliana-Alexandra, 153
Fleischer, Daniel, 154
Flekova, Lucie, 16, 177
Flor, Michael, 11, 12, 46, 67, 171, 181
Florian, Radu, 12
Floyd, Sammy, 18
Fokkens, Antske, 161
Földiák, Péter, 161
Fomin, Vadim, 182

Fone, James, 49
Fooths, Ashley, 18
François, Thomas, 170
Francesco Lotito, Quintino, 17
Frank, Anette, 11, 64
Frank, Robert, 97
Fraser, Alexander, 59
Frassineti, Diego, 165
Frermann, Lea, 140
Fried, Daniel, 136
Friedenberg, Meir, 174
Fry, Ben, 127
Fu, Mingming, 161
Fu, Yao, 44
Fujimoto, Scott, 135
Fukayama, Satoru, 43
Fung, Pascale, 154
Funke, Andreas, 163
Fürnkranz, Johannes, 133
Fyshe, Alona, 146, 183

Gábor, Kata, 159
Gaddy, David, 92
Gajananan, Kugamoorthy, 161
Gajjam, Jayashree Aanand, 170
Galbraith, Byron, 161
Galescu, Lucian, 186
Gallagher, Ryan, 54
Galley, Michel, 11
Gamallo, Pablo, 162
Gambäck, Björn, 161
Ganesh H. B., Barathi, 155, 157
Ganesh, Ananya, 180
Ganeshkumar, Premkumar, 81
Gangadharaiah, Rashmi, 40
Ganguly, Debasis, 48, 147
Ganu, Hrishikesh, 179
Gao, Fei, 85
Gao, Guang, 144
Gao, Jianfeng, 44
Gao, Sheng, 44
Gao, Yanjun, 171
Gao, Yimei, 34
Gao, Zi Yuan, 154
Gardent, Claire, 11, 15, 22
Gardner, Matt, 150
Gargett, Andrew, 165
Garten, Justin, 12
Gasic, Milica, 141
Gaspers, Judith, 78
Ge, Tao, 12
Ge, Xie, 163
Gebreselassie, Tewodros, 18
Gedeon, Tom, 156
Gee, Grace, 155
Gelbukh, Alexander, 171
-

- Gella, Spandana, 17, 144
 Gencheva, Pepa, 65
 Genthal, Guillaume, 67
 George, Anon, 155
 Georges-Filteau, Jeremy, 135
 Georgila, Kallirroi, 11
 Gerber, Ben, 19
 Gerber, Luciano, 157
 Gero, Katy, 181
 Ghaeini, Reza, 116
 Ghaffari, Parsa, 65
 Ghanem, Bilal, 157
 Ghanimifard, Mehdi, 181, 186
 Ghazi, Wegdan, 157
 Ghazvininejad, Marjan, 44, 104, 178
 Ghosh, Aniruddha, 157
 Ghosh, Debanjan, 11, 12
 Gifu, Daniela, 153, 156, 161
 Gildea, Daniel, 125
 Gilead, Michael, 175
 Giles, C Lee, 171
 Gillick, Jon, 36, 178
 Gilroy, Sorcha, 150
 Gimenez, Jesus, 36
 Gimpel, Kevin, 35, 122, 143, 167
 Ginés i Ametllé, Joan, 154
 Giorgi, Salvatore, 175
 Girju, Roxana, 155
 Gispert, Adrià de, 71, 104
 Gjurković, Matej, 184
 Gkotsis, George, 174
 Glass, James, 42, 84, 105, 125, 144
 Glasz, Christopher, 135
 Glavaš, Goran, 59, 60, 180
 Glover, John, 65
 Gluhak, Martin, 160
 Godley, Amanda, 170
 Goharian, Nazli, 129, 160, 175
 Gokcen, Ajda, 92
 Goldberg, Adele, 18
 Goldwasser, Dan, 17
 Goldwater, Sharon, 106, 142
 Gómez-Rodríguez, Carlos, 57, 88, 90, 138, 184
 Gong, Hongyu, 89
 González, Fabio A., 63, 99, 107
 González, José-Ángel, 157, 162, 163
 Gonzalez-Garduño, Ana Valeria, 140
 Goo, Chih-Wen, 144
 Gooding, Sian, 170
 Goodman, Alissa, 174
 Goodman, Noah, 81
 Gopidi, Amitha, 18
 Gordon, Andrew, 178
 Gordon, Carla, 18
 Gorinski, Philip John, 132
 Goto, Masataka, 43
 Gouveia da Silva, Thiago, 180
 Gowda, Prashanth, 155
 Goyal, Anuj Kumar, 78
 Goyal, Pawan, 50, 57
 Goyal, Tanya, 183
 Grabmair, Matthias, 168
 Graff, Mario, 153
 Graham, Yvette, 11
 Grangier, David, 40, 48
 Grave, Edouard, 98
 Gravier, Christophe, 45, 131
 Greenberg, Clayton, 81
 Gridach, Mourad, 19
 Grigoriev, Dmitry, 157
 Grishin, Maxim, 162
 Grishina, Yulia, 185
 Groeneveld, Dirk, 49
 Groot, Daphne, 156
 Grundkiewicz, Roman, 67, 68
 Grune, Barbara, 160
 Grusky, Max, 77
 Gu, Gen, 34
 Gu, Jiatao, 12, 39
 Guha, Shubha, 67
 Guibon, Gaël, 157
 Guide, Bruno, 18
 Gulordava, Kristina, 98
 Guntuku, Sharath Chandra, 175
 Guo, Lei, 18
 Guo, Linsen, 131
 Guo, Xiaoxiao, 100
 Guobin, Sui, 164
 Gupta, Ashim, 181
 Gupta, Itika, 19
 Gupta, Manish, 18
 Gupta, Pankaj, 34, 95
 Gupta, Prakhar, 61
 Gupta, Raj Kumar, 154
 Gur, Izzeddin, 85
 Gurdasani, Amit, 122
 Gureckis, Todd, 170
 Gurevych, Iryna, 42, 55, 65, 95, 116, 131, 136,
 159, 165, 167
 Gururangan, Suchin, 46
 Gustafson, Erin, 169
 Guzman, Francisco, 11
 Gyawali, Binod, 67
 Ha, Le An, 162, 172
 Ha, Thi-Thanh, 18
 Ha, Vu, 49
 Haagsma, Hessel, 156
 Habash, Nizar, 90
 Habernal, Ivan, 55, 136, 159, 176
 Hachey, Ben, 175
 Haddad, Hatem, 17, 19, 153

- Haddow, Barry, 102
Haffari, Gholamreza, 99, 104
Hagiwara, Masato, 169
Hahn, Udo, 135
Hahn-Powell, Gus, 180
Haji Soleimani, Behrouz, 155
Hajishirzi, Hannaneh, 87, 139, 140, 142–144,
 160
Hajj, Hazem, 154, 166
Hakimi Parizi, Ali, 162
Hakkani-Tür, Dilek, 35, 40, 114, 141
Haldar, Aparajita, 167
Haldar, Rajarshi, 50
Hale, John, 185
Hall, Keith, 12
Halpern, David, 170
Hamon, Thierry, 11
Hamza, Wael, 125
Han, Jiawei, 179
Han, Julien, 175
Han, Xianpei, 12, 84
Handler, Abram, 132
Hangya, Viktor, 59
Hao, Shudong, 95
Hardalov, Momchil, 157
Hardmeier, Christian, 185
Hardt, Daniel, 61
Hare, Jonathon, 131
Hartmann, Nathan, 171
Harzallah, Mounira, 161
Hasan, Sadid A., 116
Hasanuzzaman, Mohammed, 76
Hasler, Eva, 71, 104
Hassan, Arshia Zernab, 161
Hassan, Hany, 39
Hastie, Helen, 11
Hathi, Shobhit, 127
Häfty, Anna, 83
Hauff, Claudia, 171
Hayashi, Katsuhiko, 60, 130
Hazarika, Devamanyu, 63, 143
He, Di, 103
He, He, 122
He, Luheng, 124, 137
He, Xiaodong, 44, 101, 129, 143
He, Yulan, 64, 131
Heafield, Kenneth, 67, 70
Heck, Larry, 141
Heilman, Michael, 11
Heinzerling, Benjamin, 12
Helms, Casper Veistrup, 155
Henderson, John, 163
Handler, James, 11
Henry, Cassidy, 18
Hepple, Mark, 18, 69
Hercig, Tomáš, 153, 157, 161
Hernández Farfás, Delia Irazú, 158
Herzig, Jonathan, 120
Hessel, Jack, 146
Hettinger, Lena, 160
Higgins, Derrick, 11
Hill, Ethan, 92
Himeno, Shinnosuke, 153
Hirao, Tsutomu, 130–132
Hiray, Sushant, 152
Hirschberg, Julia, 136, 176
Hobbs, Reginald, 178
Hollenstein, Nora, 159
Hollingshead Seitz, Kristy, 76
Holtzman, Ari, 148
Holz, Heiko, 170
Homan, Christopher, 140
Honda, Ukyo, 132
Hong, Xudong, 165
Hongdou, Yao, 163
Hopkinson, Andrew, 122
Horbach, Andrea, 172
Hori, Chiori, 11
Hosseini, Seyedarian, 168
Hoste, Veronique, 11, 152, 153
Hotho, Andreas, 160
Hou, Yufang, 41, 133
Houben, Geert-Jan, 171
Hough, Alfred, 180
Hough, Julian, 11
Hovy, Dirk, 98, 173, 183
Hovy, Eduard, 128, 180
Hsieh, Cho-Jui, 80
Hsu, Keng-Wei, 144
Hsu, Yun-Kai, 144
Htut, Phu Mon, 126
Hu, Baotian, 45
Hu, Hai, 166, 177
Hu, Hexiang, 56
Hu, Jennifer, 145
Huang, Chenyang, 43
Huang, Hen-Hsen, 11, 162
Huang, Heyan, 155
Huang, Kenneth, 178
Huang, Laetitia, 108
Huang, Lifu, 126, 170
Huang, Po-Sen, 44, 143
Huang, Qiuyuan, 101
Huang, Ruihong, 42, 85
Huang, Shujian, 104
Huang, Yongfeng, 152, 154, 156, 162, 182
Huet, Stéphane, 180
Hulden, Mans, 108
Hulub, Ionut, 161
Huo, Chih-Li, 144
Hurtado, Lluís-F., 157, 162, 163
Huss, Anders, 170

- Hussein, Amad, 169
 Husseini Orabi, Ahmed, 154, 175
 Husseini Orabi, Mahmoud, 154, 175
 Hwa, Rebecca, 11, 60
 Hwang, Seung-won, 77
- Idiart, Marco, 18, 61
 Iglesias, Gonzalo, 71, 104
 Illic, Suzana, 157
 Ilievski, Filip, 152
 Inkpen, Diana, 154, 175
 Inui, Kentaro, 43, 100
 Ionescu, Radu Tudor, 170
 Ireland, Molly, 174, 175
 Irsoy, Ozan, 84
 Iserman, Micah, 174, 175
 Ishihara, Takahiro, 60
 Issa Alaa Aldine, Ahmad, 161
 Issa, Fuad, 57
 Iter, Dan, 175
 Ive, Julia, 174
 Iyer, Srinivasan, 150
 Iyyer, Mohit, 122, 145, 150
- J R, Naveen, 155, 157
 Jaakkola, Tommi, 54
 Jabreel, Mohammed, 154
 Jabri, Allan, 56
 Jaech, Aaron, 127
 Jäger, Gerhard, 88
 Jaggi, Martin, 61
 Jahan, Labiba, 16
 Jain, Parag, 118, 129
 Jain, Royal, 152
 Jalal, Mona, 18
 Jana, Abhik, 57
 Janicka, Maria, 182
 Jansen, Peter, 180
 Jaradat, Israa, 65
 Jastrzebski, Stanislaw, 168
 Jeon, Byungsoo, 41
 Jha, Rahul, 78
 Ji, Heng, 4, 92, 114, 170
 Ji, Yangfeng, 150
 Jia, Robin, 122
 Jian, Su, 16
 Jiang, Chao, 80
 Jiang, Jyun-Yu, 120
 Jiang, Nan, 142
 Jiang, Xin, 121
 Jiang, Youxuan, 129
 Jiang, Zhengping, 163
 Jiang, Zhiying, 170
 Jiménez-Zafra, Salud Marfa, 153
 Jin, Di, 160
 Jin, Hongxia, 83
- Jin, Lifeng, 137, 169
 Jin, Mingmin, 126
 Jin, Qiao, 42
 Jin, Shuning, 156
 Jindal, Vasu, 147
 Jo, Yohan, 41
 Jochim, Charles, 133
 Johnson, Kristen, 17
 Johnson, Mark, 92
 Jones, Gareth, 48
 Jones, Michael, 76
 Jong, Andrew, 145
 Joshi, Anirudh, 164
 Joshi, Anupam, 161
 Jovanovic, Dejan, 178
 Ju, Meizhi, 116
 Junczys-Dowmunt, Marcin, 11, 67, 68
 Jundi, Hussein, 166
 Jung, Kyomin, 125, 153
 Jurafsky, Dan, 67, 127, 175
 Juraska, Juraj, 43
 Jurczyk, Tomasz, 140
- K M, Annervaz, 38
 K P, Soman, 155, 157, 162
 K, Shalini, 155
 Kaffee, Lucie-Aimée, 131
 Kahse, Tobias, 42
 Kain, Raslan, 154
 Kajiwara, Tomoyuki, 106, 170, 171
 Kale, David, 54
 Kaljahi, Rasoul, 11, 12
 Kallmeyer, Laura, 166
 Kalouli, Aikaterini-Lida, 166, 176
 Kalousis, Alexandros, 85
 Kamath, Aishwarya, 54
 Kamigaito, Hidetaka, 130
 Kamila, Sabyasachi, 76
 Kanagasabai, Nirmal, 135
 Kaneko, Masahiro, 171
 Kang, Dongyeop, 128
 Kang, Woo Hyun, 157
 Kang, Yiping, 40
 Kann, Katharina, 16, 35
 Kao, Anne, 82
 Kao, Hung-Yu, 163
 Kar, Sudipta, 63
 Karagiannis, Panagiotis, 43
 Karanasou, Penny, 78
 Karasalo, Maja, 153
 Karlekar, Sweta, 17, 139
 Karttunen, Lauri, 11
 Karyakin, Yuri, 157
 Kasai, Jungo, 91, 97
 Kashefi, Omid, 60
 Kasunic, Anna, 178

- Katiyar, Arzoo, 87
Kato, Tsuneaki, 96
Katz, Boris, 139
Kauchak, David, 11
Kaufman, Geoff, 178
Kaur, Barleen, 135
Kavuluru, Ramakanth, 142, 175
Kawahara, Daisuke, 62
Kayser, Michael, 82
Kazantseva, Anna, 11
Kehler, Andrew, 11
Keiper, Thorsten, 160
Keith, Katherine, 90
Kelleher, John, 186
Keller, Frank, 144
Kenyon-Dean, Kian, 11, 135, 165
Khac Linh, Pham, 17
Khadaj, Alaa, 154
Khadivi, Shahram, 70
Khapra, Mitesh M., 118, 129
Khashabi, Daniel, 47
Kiela, Douwe, 56
Kiely, Kim, 175
Kim, Dongchan, 36
Kim, Doo Soon, 129
Kim, Gunhee, 120
Kim, Joo-Kyung, 36
Kim, Nam Soo, 157
Kim, Seokhwan, 129
Kim, Sujeong, 147
Kim, Taeuk, 163, 166
Kim, Yanghoon, 153
Kim, Young-Bum, 36
Kim, Youngmin, 155
King, David, 169
King, Levi, 170
King, Milton, 162
Kinney, Rodney, 49
Kiritchenko, Svetlana, 152, 155, 165, 167, 173
Kiritsis, Dimitris, 85
Kirov, Christo, 107
Klamí, Arto, 64
Klein, Dan, 92, 136, 145
Klerke, Sigrid, 169, 170
Klie, Jan-Christoph, 116
Klubička, Filip, 161
Klymenko, Momo, 141
Knight, Kevin, 44, 74, 92, 104, 150, 178
Knott, Alistair, 11
Knowles, Rebecca, 18
Kobayashi, Hayato, 11
Kobayashi, Sosuke, 100
Kochmar, Ekaterina, 169, 170
Koehn, Philipp, 12, 18
Kohlmeier, Sebastian, 49, 128
Kojaku, Sadamori, 167
Kojima, Noriyuki, 146
Kolhatkar, Varada, 17, 185
Kollar, Thomas, 82
Koller, Alexander, 11
Kolovou, Athanasia, 154, 156, 158
Kolyvakis, Prodromos, 85
Komachi, Mamoru, 11, 106, 170, 171
Konkol, Michal, 161
Konopnicki, David, 120
Köper, Maximilian, 57, 61, 165, 185
Kopev, Daniel, 157
Koppel, Moshe, 177
Kopru, Selcuk, 36
Kordjamshidi, Parisa, 147, 186
Korhonen, Anna, 60
Koshorek, Omri, 100
Kotakonda, Bhaskar, 155
Koychev, Ivan, 157
Kraft, Bodo, 49
Kravchenko, Dmitry, 153
Krebs, Alicia, 159
Krestel, Ralf, 122
Kreutzer, Julia, 70
Krishna, Kundan, 130
Krishnamurthy, Gangeshwar, 63
Kriz, Reno, 45
Kruizinga, Rémon, 156
Kübler, Sandra, 177
Kuculo, Tin, 164
Kuhn, Jonas, 77
Kuhnle, Alexander, 168
Kuijper, Marloes, 154
Kulkarni, Chaitanya, 46
Kulkarni, Mayank, 171
Kulkarni, Vivek, 107
Kulmizev, Artur, 154, 162
Kulshreshtha, Devang, 158
Kumar M, Anand, 155, 157, 162
Kumar Singh, Anil, 158
Kumar, Abhishek, 62
Kumar, Vishwajeet, 92
Kummerfeld, Jonathan K., 12, 40, 129, 143
Kuriakose, John, 130
Kurohashi, Sadao, 62
Kurosawa, Michiki, 106
Kusumawardani, Renny Pradina, 16
- Lacroix, Ophélie, 89
Laforest, Frederique, 45, 131
Laha, Anirban, 118, 129
Lai, Sunny, 159
Lai, Tuan, 99
Lakew, Surafel Melaku, 17
Lalande, Auguste, 135
Lall, Brejesh, 155
LaMar, Michelle, 170

- Lambert, Patrik, 11
Lan, Man, 154, 156, 158, 162, 163
Lan, Wuwei, 58
Laparra, Egoitz, 152
Lapata, Mirella, 11, 38, 55, 117, 132
Lappin, Shalom, 181
Lapresi, Luke, 140
Lasecki, Walter, 129
Lassoued, Yassine, 133
Lauscher, Anne, 160
Lavid-López, Julia, 183
Lavrencic, Louise, 175
Lavrentovich, Alexandra, 69
Le, Ngoc Tan, 17
Leacock, Claudia, 169
Lee, Changki, 158
Lee, Hung-yi, 125
Lee, Hwanhee, 153
Lee, Hyunju, 155, 159
Lee, John, 11
Lee, Kathy, 116
Lee, Kenton, 137, 150
Lee, Lillian, 90, 110, 146
Lee, Sang-goo, 163, 166
Lefever, Els, 152
Leidner, Jochen L., 57, 83
Lenc, Ladislav, 153
Lenci, Alessandro, 159, 165
Lenthe, Mike van, 154
Leonard, Brian, 41
Leong, Chee Wee (Ben), 46, 181
Leoznov, Artem, 157
Lepage, Yves, 11
Lester, James, 11
Leung, Kwong Sak, 159
Leung, Yee, 159
Leusch, Gregor, 45, 82
Levacher, Killian, 133, 147
Levine, Sergey, 145
Levitana, Rivka, 76, 174
Levitana, Sarah Ita, 136
Levow, Gina-Anne, 179
Levy, Omer, 11, 46, 168
Levy, Roger, 139
Li, Bin, 61
Li, Binyang, 154
Li, Bofang, 179
Li, Chen, 159
Li, Chenliang, 44
Li, Guanlin, 130
Li, Jiangtong, 161
Li, Jing, 55, 58, 129
Li, Jixing, 185
Li, Juncen, 122
Li, Lei, 100
Li, Linlin, 159, 161
Li, Meng, 154
Li, Mengping, 49
Li, Mu, 130
Li, Quanzhi, 158
Li, Ruizhe, 162
Li, Si, 44
Li, Sujian, 44
Li, Victor O.K., 39
Li, Wei, 44
Li, Wen, 177
Li, Wenjie, 44
Li, Xiang, 12
Li, Xiaolong, 127
Li, Xilian, 156
Li, Xintong, 103
Li, Yanran, 11
Li, Yitong, 101
Li, Yongbin, 163, 164
Li, Yunyao, 6, 49, 52
Li, Zifan, 126
Liang, Chao-Chun, 68
Liang, Chen, 171
Liang, He, 156
Liang, Percy, 122, 168
Liang, Yiyun, 108
Liao, Quanlei, 163
Liceralde, Van Rynald T., 144
Lichtenstein, Patricia, 181
Liebeck, Matthias, 163
Lim, Seonjae, 77
Lin, Chenghua, 162
Lin, Chu-Cheng, 11, 90
Lin, Jimmy, 49, 68, 108
Lin, Lei, 163
Lin, Yi-Chung, 68
Lin, Ying, 2, 92
Ling, Wang, 179
Ling, Yuan, 116
Linhares Pontes, Elvys, 180
Linhares, Andréa Carneiro, 180
Lino, Kenny W., 156
Linzen, Tal, 98
Lippincott, Tom, 183
List, Johann-Mattis, 88
Litman, Diane, 170–172
Littlefield, Andrew, 174
Litvinova, Olga, 175
Litvinova, Tatiana, 175
Liu, Bing, 11, 12, 19, 40, 69, 141
Liu, Bingquan, 142, 163
Liu, Chenxi, 55
Liu, Chunhua, 163
Liu, Fei, 64
Liu, Frederick, 103
Liu, Honglei, 85
Liu, Jiangming, 63

- Liu, Jingming, 159
Liu, Joey, 116
Liu, Junxin, 152, 154, 156
Liu, Lemao, 103
Liu, Liang-Chih, 133
Liu, Lizhen, 164
Liu, Lu, 163
Liu, Man, 155
Liu, Nelson F., 179
Liu, Qi, 63, 142
Liu, Qun, 35
Liu, Shujie, 124, 130
Liu, Tao, 179
Liu, Tie-Yan, 85, 103
Liu, Ting, 63
Liu, Wenjie, 163
Liu, Xuan, 143
Liu, Yang, 38, 67
Liu, Yijia, 91
Liu, Yingchi, 158
Liu, Yu-Wen, 133
Liu, Yuhan, 154
Livescu, Karen, 11, 35
Lloret, Elena, 11, 12
Lo, Kai-Ling, 44
Lo, Kyle, 49
Loáiciga, Sharid, 185
Lockard, Colin, 87
Lønning, Jan Tore, 160
Loos, Andreas, 122
Lopes, Jose David, 12
Lopez Condori, Roque, 167
López Monroy, Adrian Pastor, 99, 107
Lopez, Adam, 97
Lorent, Alina, 156
Louis, Annie, 141
Loukina, Anastassia, 48, 144, 169
Lovelett, Jarrett, 170
Loveys, Kate, 174, 175
Lowry-Duda, Joanna, 162
Loyola, Pablo, 161
Loza Mencía, Eneldo, 133
Lu, Di, 12, 92
Lu, Han, 103
Lu, Wei, 159
Lu, Xingwu, 156
Lu, Zhiyong, 122
Luan, Yi, 160
Lucas, Andrew T., 60
Lucic, Ivana, 171
Lugini, Luca, 170
Luh, Wen-Ming, 185
Lukin, Stephanie, 11, 18, 178
Luo, Fan, 180
Luo, Xin, 126
Luo, Zhunchen, 121, 160
Luong, Daphne, 186
Lv, Yuanhua, 124
Lynn, Veronica, 174
Lyu, Zhonghao, 160
Ma, Chunping, 159
Ma, Kaixin, 140
Ma, Shuming, 44
Maarouf, Alaa, 154
MacAvaney, Sean, 62, 160, 175
Machiraju, Raghu, 46
Macke, Stephen, 179
Madgula, Krishna, 161
Madhyastha, Pranava Swaroop, 81, 145
Madnani, Nitin, 11, 48, 169
Maffey, Luke, 153
Mager Hois, Jesus Manuel, 35
Mager, Manuel, 16
Mahabal, Abhijit, 167
Mahajan, Praateek, 183
Maharjan, Suraj, 63
Mahata, Debanjan, 130
Mahendra, Rahmad, 152
Mahendran, Darshini, 160
Maheshwari, Ayush, 92
Maier, Wolfgang, 162
Makarucha, Adam, 11
Makrai, Márton, 161
Malandrakis, Nikolaos, 12
Maldonado, Alfredo, 161
Male, John, 65
Maletti, Andreas, 150
Maliepaard, Sage, 174
Malliaros, Fragkiskos, 180
Malmasi, Shervin, 169, 171
Malykh, Valentin, 17
Mamdouh, Sherine, 157
Mamidi, Radhika, 18
Manabe, Hitoshi, 60
Manjunatha, Varun, 145
Manning, Christopher D., 97
Mansfield, Courtney, 17
Manzoor, Umar, 147, 186
Mao, Rui, 162
Mao, Xin, 156
Maoquan, Wang, 156
Marasovic, Ana, 64, 185
Marcheggiani, Diego, 105
Marciniak, Małgorzata, 181, 182
Marcu, Daniel, 39, 93
Maredia, Angel, 136
Marge, Matthew, 18
Marin, Alex, 78
Márquez, Lluís, 42, 65, 84
Marrese-Taylor, Edison, 157
Mars, Jason, 40
-

- Marshall, Iain, 88
Martin, Maite, 153
Martindale, Marianna, 17
Martínez Alonso, Héctor, 12, 170
Martinez, Ander, 141
Martínez-Gómez, Pascual, 84
Martyn-Nemeth, Pamela, 70
Marvin, Rebecca, 184
Masuda, Naoki, 167
Mathias, Lambert, 82
Mathias, Sandeep, 170
Mathur, Kritik, 18
Mathur, Prashant, 82
Mathur, Vinayak, 180
Matsoukas, Spyros, 78, 82
Matsabayashi, Yuichiroh, 43
Matsumura, Yukio, 106
Matsuo, Yutaka, 157
Matsuyama, Yoichi, 11
Matthews, Austin, 107
Matusov, Evgeny, 70
Matwin, Stan, 155
May, Jonathan, 2, 77, 92, 150, 152, 178
Mayank Srivastava, Muktabh, 158
Mayorov, Vladimir, 160
Mays, Kate K., 18
Mazidi, Karen, 11
McCallum, Andrew, 54, 59, 87, 180
McCarthy, Diana, 11
McCrae, John Philip, 181
McCulla, Laura, 67
McDermott, Matthew, 160
McInnes, Bridget, 153, 160
McMasters, James, 127
McMillan-Major, Angelina, 17
M.Davies, Patricia, 171
Meéndez-Cruz, Carlos-Francisco, 162
Meerkamp, Philipp, 11
Mehrabani, Afshin, 65
Mehrabani, Mahnoosh, 102
Mei, Hongyan, 166
Meisherri, Hardik, 155
Melamud, Oren, 11
Melese, Michael, 16
Mencarini, Letizia, 183
Meng, Kongming, 154
Meng, Max, 103
Menzel, Wolfgang, 165
Merghani, Taha, 177
Merkhofer, Elizabeth, 163
Merrill, William, 97
Meshesha, Million, 16
Metallinou, Angeliki, 78
Meurers, Detmar, 41, 169, 170
Meza Ruiz, Ivan Vladimir, 16, 17, 35, 162
Michael, Julian, 87, 124
Miculicich Werlen, Lesly, 102
Mielke, Sebastian J., 107
Mieskes, Margot, 17
Miftahutdinov, Zulfat, 17
Mihalcea, Rada, 76, 143, 146, 177
Mikhalkova, Elena, 157
Miller, Mark, 49, 71
Miller, Timothy, 137, 139, 166
Miller, Tristan, 65
Milne, David, 174
Miltsakaki, Eleni, 45
Mimmo, David, 80, 146
Min, Bonan, 116
Minervini, Pasquale, 168
Mineshima, Koji, 84, 89
Miranda-Jiménez, Sabino, 153
Mirza, Madiha, 157
Mirza, Paramita, 152
Misra, Ishan, 178
Mitchell, Jeff, 12, 138, 168
Mitchell, Margaret, 55, 173, 178
Mitcheltree, Christopher, 48
Mitkov, Ruslan, 157, 162
Mitra, Pabitra, 181
Mittal, Arpit, 85, 122
Mittal, Sid, 167
Miwa, Makoto, 116
Miyao, Yusuke, 121
Moctezuma, Daniela, 153
Modi, Ashutosh, 159
Moens, Marie-Francine, 108, 186
Mohammad, Saif, 11, 12, 152, 155, 165, 167,
 173
Mohammed, Salman, 68
Mohan, Mahesh, 62
Mohtarami, Mitra, 42, 84
Mokhtari, Shekoofeh, 18
Möller, Verena, 170
Molloy, Hillary, 67
Monroe, Will, 145
Montes, Manuel, 63, 99, 158
Moon, Seungwhan, 86
Moon, Taesun, 11
Mooney, Raymond, 38, 146
Moosavi, Nafise, 185
Mor, Noam, 100
Mora, Natalia, 183
Morales, Michelle, 174
Morante, Roser, 176
Moreau, Anatole, 153
Morency, Louis-Philippe, 143, 146
Moreno, Antonio, 154
Morgan, Nicholas, 153
Moriarty, Glen, 175
Moschitti, Alessandro, 42, 84
Mosolova, Anna, 182

- Moss, Larry, 166
Mostafazadeh, Nasrin, 11
Mostafazede, Nasrin, 178
Mousselly Sergieh, Hatem, 116, 167
Mozgai, Sharon, 174
Mrkšić, Nikola, 15, 27, 28, 60, 96
Mukherjee, Animesh, 50
Mulholand, Matthew, 12
Mulki, Hala, 17, 153
Mullenbach, James, 95
Muller, Philippe, 11
Munkhdalai, Tsendsuren, 45
Muresan, Smaranda, 135, 181
Murray, Tyler, 49
Murrieta Bello, Héctor Ricardo, 17
Mwebaze, Ernest, 16
Mykowiecka, Agnieszka, 181, 182

Naaman, Mor, 77
Nadeem, Farah, 169
Naderi, Habibeh, 155
Nagata, Masaaki, 60, 130–132
Nagesh, Ajay, 86
Nagoudi, El Moatez Billah, 155
Nair, Suraj, 174
Nakamura, Satoshi, 103
Nakano, Tomoyasu, 43
Nakashole, Ndapa, 168
Nakov, Preslav, 42, 65, 84
Nandutu, Irene, 16
Nangia, Nikita, 96, 101
Naradowsky, Jason, 41, 167
Narasimhan, Karthik, 11
Narayan, Shashi, 15, 22, 132
Narayanan, Shrikanth, 154
Narayanan, Sriniv, 186
Narayanaswamy, Balakrishnan, 40
Narwekar, Abhishek Avinash, 155
Nastase, Vivi, 12
Nath, J. Saketha, 92
Naumann, Daniela, 165
Navigli, Roberto, 159
Nayak, Nihal V., 170
Nedoluzhko, Anna, 185
Negri, Matteo, 17
Nema, Preksha, 118, 129
Nematzadeh, Aida, 18
Nenková, Ani, 88
Neubig, Graham, 15, 21, 41, 101, 103, 104, 107, 179
Neumann, Günter, 160
Neumann, Mark, 150
Neves, Leonardo, 86
Neves, Mariana, 160
Ng, Andrew, 67
Ng, Vincent, 136, 185

Ngo, Thi Lan, 17
Nguyen, Dai Quoc, 83, 92, 157
Nguyen, Dat Quoc, 83, 92, 157
Nguyen, Dong, 11, 12, 95
Nguyen, Hong Buu Long, 17
Nguyen, Kiem-Hieu, 18
Nguyen, Kim Anh, 59, 185
Nguyen, Thanh-Chinh, 18
Nguyen, Thienn Huu, 12
Nguyen, Toan, 39
Nguyen, Tu Dinh, 83
Nias, Jaye, 17
Nickel, Maximilian, 56
Nicolai, Garrett, 11
Nie, Jian-Yun, 155
Niebler, Thomas, 160
Niederhoffer, Kate, 174
Nielsen, Rodney, 176
Nikhil, Nishant, 158
Nikolaos, Athanasiou, 154, 156, 158
Nikolova, Ivelina, 157
Nilsson, Mattias, 153
Nilsson, Susanna, 170
Ning, Qiang, 86, 166
Niraula, Nobal Bikram, 82
Nishino, Masaaki, 131
Nissim, Malvina, 162, 165, 183
Niu, Tong, 139
Niu, Xing, 117
Niven, Timothy, 163
Nivre, Joakim, 97
Noji, Hiroshi, 89
Nooralahzadeh, Farhad, 160
Noord, Rik van, 154
Norouzi, Mohammad, 180
Noukhovitch, Michael, 168
Novák, Michal, 185
Novikova, Jekaterina, 45
Nowson, Scott, 175
Nugent, Timothy, 57, 83
Nuxoll, Florian, 170
Nyberg, Eric, 168
Nyegaard-Signori, Thomas, 155

Obeidat, Rasha, 18
Ochs, Magalie, 157
O'Connor, Brendan, 90, 132
Ogrodniczuk, Maciej, 185
Oladiop, Francisca, 17
Olex, Amy, 153
Oliveira Lima, Guilherme, 18
Oliveira, Eugénio, 158
Oliveira, Hugo Gonçalo, 11, 12
Olsen, Mari, 186
Olshefski, Christopher, 170
Onofrei, Mihaela, 161

- Onyenwe, Ikechukwu, 18, 69
Onysko, Alexander, 182
Ooi, Hsu-Han, 49
Oraby, Shereen, 12, 18, 69
Oramas, Sergio, 159
Ordonez, Vicente, 42
Osborn Popp, Pamela, 170
Osika, Anton, 170
Ositelu, Oladipo, 161
Ostendorf, Mari, 35, 52, 127, 148, 160, 169
Ostermann, Simon, 159
Oswal, Ankita, 169
Ott, Myle, 40
Øvrelid, Lilja, 11, 160
Owczarzak, Karolina, 82
- P, Viswa Datha, 179
Padia, Ankur, 161
Padmanabhan, Sarguna, 104
Padó, Sebastian, 59, 179
Paetzold, Gustavo, 169
Pagel, Janis, 185
Pagliardini, Matteo, 61
Palfrey, Dave, 122
Pallathra, Ashley A., 175
Palmer, Alexis, 11, 58
Palmer, Martha, 152, 181
Pamungkas, Endang Wahyu, 158
Pan, Shimei, 161
Pan, Xiaoman, 92
Pandey, Juhí, 175
Pang, Bo, 110, 186
Papalampidi, Pinelopi, 158
Papay, Sean, 179
Paperno, Denis, 159
Papineni, Kishore, 110
Pappas, Nikolaos, 102
Paraskevopoulos, Georgios, 154, 156, 158
Parde, Natalie, 176
Parikh, Devi, 145, 168
Paris, Cecile, 164
Parish-Morris, Julia, 175
Park, Cheoneum, 158
Park, Ji Ho, 154
Park, Yookoon, 120
Park, Youngja, 161
Parsons, Simon, 42
Parupalli, Sreekavitha, 18
Pasini, Tommaso, 159
Pasquer, Caroline, 91
Passban, Peyman, 35
Passonneau, Rebecca J., 20, 171
Pasunuru, Ramakanth, 131
Patel, Roma, 88
Patro, Sohan, 50
Patti, Viviana, 152, 158, 183
- Paul, Michael J., 95
Paula, Felipe, 18, 61
Paullada, Amandalynne, 16
Pavalanathan, Umashanthi, 69
Pavlick, Ellie, 152, 165
Pedersen, Ted, 156, 157, 161
Peikos, Georgios, 156
Peng, Bo, 158
Peng, Hao, 117
Peng, Haoruo, 86
Peng, Nanyun, 15, 23, 80, 178
Pennebaker, James W., 177
Pereira, Fernando, 186
Perera, Ian, 186
Pérez Alvarez, Daniel Alejandro, 171
Perez-Beltrachini, Laura, 11, 41, 43, 46, 117
Peters, Matthew, 49, 150
Peters, Ruth, 175
Petroni, Fabio, 57, 83
Petrucci, Miriam R L, 186
Peyrard, Maxime, 131
Pezzelle, Sandro, 56
Pham, Khiem, 180
Phandi, Peter, 159
Phillips, Lawrence, 127
Phung, Dinh, 83
Piergallini, Mario, 17
Pierrejean, Benedicte, 62, 166
Pilán, Ildikó, 171
Pilehvar, Mohammad Taher, 15, 25
Pinkal, Manfred, 159
Pinter, Yuval, 55
Pinto, Joel, 36
Piorkowski, David, 120
Pivovarova, Lidia, 64, 153
Piwek, Paul, 11
Pla, Ferran, 157, 162, 163
Plachouras, Vassilis, 57, 83
Plank, Barbara, 19, 34, 170, 177, 179, 183, 184
Plate, Samantha, 175
Platonov, Georgiy, 186
Plaza del Arco, Flor Miriam, 153
Poddar, Shivani, 41
Poesio, Massimo, 185
Poibeau, Thierry, 11
Poliak, Adam, 105, 167
Pollard, Kimberly A., 18
Ponzetto, Simone Paolo, 160
Pooladzadeh, Sara, 160
Popescu-Belis, Andrei, 102
Popović, Maja, 171
Popowich, Fred, 11
Poria, Soujanya, 63, 143
Post, Matt, 103
Postma, Marten, 152
Potamianos, Alexandros, 154, 156, 158

- Potash, Peter, 177
Potdar, Saloni, 100
Potts, Christopher, 60, 81, 145, 167
Pourdamghani, Nima, 104
Power, Joanna, 49
Power, Russell, 47
Prabhakaran, Vinodkumar, 12, 15, 26, 81
Prakash, Aaditya, 116
Pramanick, Malay, 181
Pratap, Bhanu, 161
Precup, Doina, 165
Prendinger, Helmut, 157
Preoțiuc-Pietro, Daniel, 11
Priansya, Stezar, 16
Prud'hommeaux, Emily, 174
Pruthi, Danish, 125
Pryzant, Reid, 127
Ptucha, Raymond, 140
Pursel, Bart, 171
Pustejovsky, James, 186
Přibáň, Pavel, 153
- Qadir, Ashequl, 116
QasemiZadeh, Behrang, 159, 166
Qayyum, Faiza, 158
Qi, Ye, 104
Qian, Jing, 47
Qin, Bing, 91
Qin, Long, 171
Qin, Tao, 85, 103
Qingxun, Liu, 163
Qiu, Wei, 161
Qu, Yanru, 34
Quinn, Jerry, 71
- R, Manikandan, 161
Radev, Dragomir, 91, 126
Radford, Kylie, 175
Radford, Will, 175
Radhakrishnan, Priya, 18, 121
Rahgooy, Taher, 147
Rahman, Kazi Shefaet, 84
Rajagopal, Dheeraj, 125
Rajani, Nazneen Fatema, 146
Rajaram, Subburam, 95
Rajendran, S Milton, 155, 158
Rajendran, Pavithra, 42
Raju, Kaushik, 85
Rakshit, Geetanjali, 19
Ram, Dhananjay, 102
Rama, Taraka, 88, 152, 170
Ramakrishnan, Ganesh, 92
Ramanarayanan, Vikram, 170
Rambow, Owen, 81, 82, 97
Ramesh, Sree Harsha, 106
Ramisch, Carlos, 11, 91
- Ramsay, Allan, 154
Rangel, Francisco, 157, 177
Rangwani, Harsh, 158
Ranshous, Stephen, 127
Ranzato, Marc'Aurelio, 40
Rao, Arjun R., 170
Rao, Sudha, 41
Rapoport, Ari, 76
Rasooli, Mohammad Sadegh, 39
Ravi, Sujith, 11, 12
Ravishankar, Vinit, 162
Rayson, Paul, 11
Real, Livy, 18
Reddy, Sravana, 11
Reed, Lena, 18
Rei, Marek, 11, 38
Reichart, Roi, 99
Reimers, Nils, 95
Reing, Kyle, 54
Reiss, Frederick, 49
Reiter, Ehud, 11
Reitter, David, 140
Ren, Shuo, 130
Ren, Xiang, 15, 23, 120
Ren, Xuancheng, 44, 87
Renslow, Tyler, 160
Resnik, Philip, 56, 174
Resnik, Rebecca, 174
Reznikova, Sofia, 163
Rheault, Ludovic, 11
Ribeiro, Alison, 156
Riccardi, Giuseppe, 11
Rich, Alexander, 170
Richards, John, 120
Richardson, Kyle, 77
Riedel, Sebastian, 138, 168
Riedl, Mark, 11, 46
Riedl, Martin, 128, 180
Rieser, Verena, 16, 45, 173
Riezler, Stefan, 70
Riloff, Ellen, 135
Rim, Kyeongmin, 69
Rimel, Laura, 11
Río Gayo, Iria del, 171
Riordan, Brian, 48, 171
Rios, Anthony, 142, 175
Rios, Miguel, 80
Risch, Julian, 122
Ritter, Alan, 46
Roark, Brian, 11, 107, 179
Robertson, Alexander, 142
Robinson, Amanda, 49, 71
Rodrigues, Ana, 18
Roemmel, Melissa, 11, 178
Roesiger, Ina, 185
Rohanian, Omid, 157, 162

- Rohde, Hannah, 185
Rohrbach, Marcus, 11
Rojas Barahona, Lina M., 141
Romanov, Alexey, 177
Rong, Wenge, 142
Rønning, Ola, 61
Ronzano, Francesco, 136, 152
Rooshenas, Amirmohammad, 54
Rose, Carolyn, 41
Rosell, Magnus, 153
Rosen, Eric, 11
Rosen, Zachary, 182
Rosenfeld, Alex, 58
Rosset, Sophie, 12
Rosso, Paolo, 157, 158, 177
Roth, Benjamin, 34
Roth, Dan, 47, 67, 86, 97, 132, 166–168
Roth, Michael, 47, 159
Roth, Stefan, 116, 167
Rothe, Anselm, 170
Rotman, Michael, 100
Rotsztejn, Jonathan, 159
Roturier, Johann, 11
Roukos, Salim, 110
Roussel, Adam, 185
Roy, Arpita, 161
Roy, Subhro, 97
Rozental, Alon, 154
Ruder, Sebastian, 65, 135
Rudinger, Rachel, 41, 78, 167
Rudzewitz, Björn, 170
Rumshisky, Anna, 177
Ruppenhofer, Josef, 81
Ruths, Derek, 135
Ruvini, Jean David, 36
Rwebangira, Mugizi, 17, 19
- S, Angel Deborah, 155, 158
S, Rajalakshmi, 155, 158
Sabharwal, Ashish, 98
Sachan, Devendra, 104
Sachan, Mrinmaya, 68
Sadat, Fatiha, 17
Sadoughi, Najmeh, 49, 71
Saers, Markus, 11
Sagarkar, Manasvi, 167
Saggion, Horacio, 136, 152, 159, 170
Saha, Snehashu, 161
Saha, Sriparna, 63
Sahin, Faruk, 170
Sajjad, Hassan, 105
Sakaue, Shinsaku, 131
Salakhutdinov, Ruslan, 42
Salameh, Mohammad, 152
Salant, Shimi, 124
Salle, Alexandre, 16, 179
- Salloum, Wael, 71
Salter, David, 147
Saluja, Avneesh, 48
Samardzhiev, Krasen, 165
Samih, Younes, 162
San, Aidan, 157
Sanchez, Ivan, 138
Sánchez-Vega, Fernando, 158
Sandbank, Tommy, 120
Sankaranarayanan, Karthik, 118, 129
Sankaranarayanan, Krishna Prasad, 106
Santilli, Andrea, 156
Santos, Leandro Borges dos, 171
Santus, Enrico, 159, 162
Sap, Maarten, 148
Sarikaya, Ruhi, 36
Sariyanidi, Evangelos, 175
Sarrazingendron, Roman, 135
Sasaki, Shota, 100
Satoh, Fumiko, 161
Satyapanich, Taneeya, 161
Savary, Agata, 91
Savova, Guergana, 139
Sawant, Palaash, 64
Sayeed, Asad, 165
Schamoni, Shigehiko, 100
Scheffler, Tatjana, 185
Scherer, Stefan, 174
Schiller, Benjamin, 65
Schlechtweg, Dominik, 58
Schluter, Natalie, 12, 16, 61
Schmaltz, Allen, 11, 173
Schmidt, Anna, 81
Schneider, Andrew, 65
Schneider, Jodi, 11
Schneider, Nathan, 91, 97
Schockaert, Steven, 88
Schönfelder, Gilbert, 160
Schreiber, Marc, 49
Schubert, Lenhart, 186
Schuler, William, 137
Schulte im Walde, Sabine, 57–59, 61, 83, 165, 185
Schultz, Robert T., 175
Schulz, Claudia, 42
Schumann, Anne-Kathrin, 159
Schumann, Raphael, 36
Schuster, Sebastian, 97
Schütze, Hinrich, 16, 34, 35, 95, 179
Schwartz, H. Andrew, 11, 174, 175
Schwartz, Lane, 137
Schwartz, Roy, 11, 12, 46, 128
Scott, Donia, 11
Sedghi, Hanie, 98
Segarra, Encarna, 162, 163
Segura-Bedmar, Isabel, 160

- Semmar, Nasredine, 179
Sen, Procheta, 48
Sen, Sukanta, 64
Sennrich, Rico, 102
Seredin, Pavel, 175
Sergeeva, Elena, 160
Settles, Burr, 169
Seyffarth, Esther, 61
Sha, Fei, 56
Shah, Kashif, 36
Shah, Pararth, 40, 141
Shah, Rajiv Ratn, 130
Shaikh, Samira, 155
Shallue, Christopher, 180
Shan, Jiawei, 154
Shank, Daniel, 161
Shapiro, Pamela, 179
Shardlow, Matthew, 157
Sharma, Dipti, 91
Sharp, Lisa, 19
Sharp, Rebecca, 166
Shaw, Peter, 101
She, Lanbo, 12
Shekarchi, Zahra, 18
Shekhar, Leena, 168
Shen, Jian, 34
Shen, Kelly, 127
Shen, Kewei, 159
Shen, Qinlan, 41
Shen, Yanyao, 103
Shen, Yilin, 83
Sheng, Yixuan, 163
Sherif, Myan, 157
Sheth, Amit, 63
Shetty, Shreyas, 118, 129
Shi, Peng, 68
Shi, Shuming, 58, 103
Shi, Tianze, 90
Shimanaka, Hiroki, 106
Shimbo, Masashi, 60
Shimorina, Anastasia, 17
Shin, Hagyeong, 43, 70
Shin, Joongbo, 125
Shing, Han-Chin, 174
Shirvani, Rouzbeh, 17
Shiue, Yow-Ting, 162
Shivaprasad, Suvamsh, 78
Shiyun, Chen, 156
Shmueli-Scheuer, Michal, 120
Shrivastava, Manish, 91, 128, 183
Shuai, Wu, 160
Shugars, Sarah, 55
Shutova, Ekaterina, 152, 181
Shutyi, Svetlana, 17
Shwartz, Vered, 60, 159, 166
Si, Luo, 86, 158, 159, 161
Siddharthan, Advaith, 11
Sikdar, Utpal Kumar, 161
Sil, Avirup, 68
Silberer, Carina, 152
Silva, Amila, 159
Silva, Cindy, 18
Silva, Nádia, 156
Simaan, Khalil, 80
Simard, Michel, 11
Simchon, Almog, 175
Simonjetz, Fabian, 185
Simperl, Elena, 131
Singh, Anil Kumar, 182
Singh, Jaya, 183
Singh, Mayank, 50
Singh, Navjyoti, 18, 183
Singh, Rishabh, 143
Singh, Vinay, 128, 183
Sinnott, Richard O., 164
Sirts, Kairit, 11
Skalicky, Stephen, 181
Skianis, Konstantinos, 180
Skjonsberg, Sam, 49
Skurniak, Filip, 182
Smith, David, 18, 179
Smith, Max, 146
Smith, Noah A., 46, 91, 117, 128, 148, 150, 179
Smolensky, Paul, 101
Snow, Bradford, 82
Sochenkov, Ilya, 166
Søgaard, Anders, 38, 61, 135, 140
Soldaini, Luca, 160, 175
Solorio, Thamar, 63, 99, 107, 137, 177
Somasundaran, Swapna, 67, 180
Sommerauer, Pia, 161
Son, Kilho, 147
Song, Heejun, 158
Song, Linfeng, 125
Song, Qingqing, 156
Song, Wei, 164
Song, Yan, 58, 129
Soo, Von-Wun, 47
Soricut, Radu, 117, 186
Sorodoc, Ionut-Teodor, 56, 152
Sorokin, Daniil, 165
Souza, Rodrigo, 18
Specia, Lucia, 81, 108, 145, 169
Speer, Robert, 162
Spinks, Graham, 108
Springorum, Sylvia, 165
Srinivasan, Balaji Vasan, 130
Srinivasan, Siddarth, 46
Sripada, Somayajulu, 11
Srivastava, Shashank, 132
Stab, Christian, 65
-

- Stahlhut, Chris, 65
Stanovsky, Gabriel, 87, 124
Stanton, Peyton, 2
Stathopoulos, Yiannos, 38
Stede, Manfred, 185
Steele, David, 108
Stein, Benno, 55, 136, 159
Steinberger, Josef, 161
Stemle, Egon, 182
Stemmer, Fabian, 36
Stenetrop, Pontus, 168
Stennmanns, Sebastian, 172
Stent, Amanda, 4, 12, 19, 74, 121
Stephenson, Martin, 147
Sterckx, Lucas, 175
Stern, Benjamin, 102
Stern, Mitchell, 92
Stevenson, Suzanne, 18
Stewart, Ian, 55
Stewart, Robert, 174
Stierman, Nathalie Adriënne Jacqueline, 156
Stone, Matthew, 11
Stowe, Kevin, 181
Stoyanov, Veselin, 12
Strickhart, Laura, 163
Striegnitz, Kristina, 11
Strube, Michael, 173
Strubell, Emma, 87
Stuart, Lauren, 82
Su, Keh-Yih, 12, 68
Su, Pei-Hao, 15, 27, 141
Su, Qi, 87
Su, Shang-Yu, 44, 144
Su, Yu, 85
Suárez-Paniagua, Víctor, 160
Subramanian, Shivashankar, 137
Suendermann-Oeft, David, 49, 71
Suhr, Alane, 150
Sulem, Elior, 76
Suleiman, Kaheer, 62, 70
Sumita, Eiichro, 103
Sun, Chengjie, 163
Sun, Huan, 85
Sun, Jimeng, 95
Sun, Le, 84
Sun, Meng, 159
Sun, Qi, 163
Sun, Shuo, 100
Sun, Xu, 44, 87
Sun, Yibo, 124
Suominen, Hanna, 156
Surdeanu, Mihai, 38, 86, 180
Suster, Simon, 124
Sutton, Charles, 141
Suvorov, Roman, 166
Swanberg, Kevin, 157
Swarnkar, Krishnkant, 182
Swayamdipta, Swabha, 46, 117
Sydorchuk, Andrii, 170
Sylak-Glassman, John, 11
Symeonidis, Symeon, 156
Sysoev, Andrey, 160
Szubert, Ida, 97
T T, Mirnalinee, 155, 158
Taboada, Maite, 17
Tachbelie, Martha Yifiru, 17
Tack, Anaïs, 169–171
Tadepalli, Prasad, 18
Takhanov, Rustem, 106
Talmor, Alon, 68
Talukdar, Partha, 121
Tam, Yik-Cheung, 88
Tambellini, William, 71
Tamrakar, Amir, 147
Tan, Hao, 81
Tan, Xu, 103
Tandon, Niket, 126
Tang, Bingjie, 161
Tang, Duyu, 124
Tang, Lingjia, 40
Tanguy, Ludovic, 62, 166
Taslimipoor, Shiva, 157, 162
Tauchmann, Christopher, 65
Tellez, Eric S., 153
Teng, Choh Man, 167, 186
Tesauro, Gerald, 100
Tetreault, Joel, 41, 150, 169
Teufel, Simone, 38
Thadani, Kapil, 11
Thalenberg, Bruna, 18
Thater, Stefan, 159
Theune, Mariët, 11
Thomson, David, 102
Thomson, Sam, 117
Thorpe, James, 85
Tian, Junfeng, 163
Tighe, Edward, 184
Titov, Ivan, 105
Tomoschuk, Brendan, 170
Torabi Asr, Fatemeh, 76
Toral, Antonio, 11
Torfi, Amirsina, 17
Tornblad, McKenna, 140
Torres-Moreno, Juan-Manuel, 180
Torrez, Jonathan, 175
Toshniwal, Shubham, 35
Tozzo, Alex, 178
Trabelsi, Amine, 43
Tran, Nam Khanh, 167
Tran, Quan Hung, 99
Tran, Trang, 35

- Tran, Tung, 175
Trancoso, Isabel, 179
Trandabat, Diana, 153, 156, 161
Traum, David, 18
Trenell, Michael, 157
Trischler, Adam, 62, 70
Tsai, Chen-Tse, 12
Tsai, Ming-Feng, 133
Tsai, Richard Tzong-Han, 84
Tseng, Bo-Hsiang, 141
Tseng, Michael, 186
Tsvetkov, Yulia, 15, 26, 80, 179
Tu, Lifu, 167
Tu, Zhaopeng, 103
Tu, Zhucheng, 49, 108
Tunc, Birkan, 175
Tür, Gokhan, 12, 40, 122, 141
Turchi, Marco, 11, 17
Turcu, Ramona-Andreea, 153
Tutubalina, Elena, 17

Ueffing, Nicola, 45, 82
Uiitdenbogerd, Alexandra, 177
Ulinski, Morgan, 176
Ultes, Stefan, 141
Uma, Alexandra, 185
Ungar, Lyle, 175, 183
Unnithan, Nidhin A, 155
Upadhyay, Shyam, 47, 67
Uppal, Karan, 155
Urena Lopez, L. Alfonso, 153
Uryupina, Olga, 185
Uszkoreit, Jakob, 101
Utayama, Masao, 103

Šnajder, Jan, 59, 137, 160, 164, 184
Štajner, Sanja, 169
V, Hariharan, 157
Vaithyanathan, Shivakumar, 110
Vajjala, Sowmya, 170, 171
Valenzuela-Escárcega, Marco A., 180
Vallabhajosyula, Manikya Swathi, 161
Van Bruwaene, David, 154
van Deemter, Kees, 11
Van Durme, Benjamin, 41, 78, 105, 166, 167
Van Hee, Cynthia, 152
Van Miltenburg, Emiel, 11
Varma, Vasudeva, 18, 121
Vaswani, Ashish, 101
Vazirgiannis, Michalis, 180
Veale, Tony, 157, 181
Veldthuis, Hennie, 156
Veloso, Adriano, 46
Velupillai, Sumithra, 174
Vempala, Alakananda, 58
Ver Steeg, Greg, 54

Verga, Patrick, 87
Verma, Rohit, 135
Vie, Jill-Jênn, 171
Vij, Prateek, 63
Vijay, Deepanshu, 128, 183
Vilar, David, 103, 105
Vilares, David, 57, 184
Villavicencio, Aline, 16, 18, 61, 179
Vilnis, Luke, 59
Vinayan, Vivek, 162
Viswanath, Pramod, 89
Vlachos, Andreas, 11, 85
Vogel, Stephan, 105
Voigt, Rob, 15, 26
Volkova, Svitlana, 12, 127
Volodina, Elena, 169
Voronov, Alexander, 157
Voss, Clare, 18, 81, 178, 186
Vossen, Piek, 152, 158, 161
Vougiouklis, Pavlos, 131
Vozila, Paul, 36
Vu, Ngoc Thang, 59, 179
Vu, Thanh, 92, 157
Vu, Tu, 45, 166
Vu, Van-Chung, 18
Vu, Xuan-Son, 157
Vulić, Ivan, 15, 27, 59, 60, 96
Vyas, Yogarshi, 67, 117

Wachsmuth, Henning, 55, 136, 159
Wadhwa, Soumya, 168
Wagner, Claudia, 183
Wagner, Joachim, 12
Wagner, Stefan, 127
Wahle, Johannes, 88
Walker, Marilyn, 2, 18, 32, 43, 110, 182
Wallace, Byron, 88
Walsh, Timothy, 175
Wang, Baoxun, 142
Wang, Chenglong, 143
Wang, Chi, 12
Wang, Chuan-Ju, 133
Wang, Chuan, 61
Wang, Eugene, 155
Wang, Feixiang, 158
Wang, Feng, 102
Wang, Haoyu, 100
Wang, Jin, 154, 156, 158, 163
Wang, Josiah, 81, 145
Wang, Liang, 159
Wang, Lucy, 49
Wang, Lu, 55, 124, 126, 128
Wang, Mingzhe, 146
Wang, Min, 154
Wang, Nan, 156
Wang, Su, 77

- Wang, Tengjiao, 156
Wang, Tianlu, 42
Wang, Wei, 120
Wang, Wenting, 158
Wang, William Yang, 15, 23, 47, 100, 107,
 116, 120, 147
Wang, Xiaolong, 142
Wang, Xin, 147
Wang, Yicheng, 125
Wang, Yu-Siang, 55
Wang, Yuan-Fang, 147
Wang, Yu, 83
Wang, Zhenduo, 156, 157
Wang, Zhenghui, 34
Wang, Zhiguo, 125
Wang, Zhuoran, 142
Wang, Ziyun, 59
Wani, Nikhil, 170
Wanner, Leo, 11
Ward, Todd, 110
Waseem, Zeerak, 16
Washio, Koki, 96
Watanabe, Kento, 43
Watanabe, Taro, 12
Watanabe, Yuji, 161
Waterson, Chris, 12, 60
Wawer, Aleksander, 181, 182
Way, Andy, 35, 76
Weber, Noah, 168
Wee, Chee, 181
Wei, Luo, 160
Wei, Ran, 155
Weise, Andreas, 76
Weiss, Zarah, 170
Weld, Daniel, 87
Wen, Ji, 87
Wendlandt, Laura, 143
Wenhan, Chao, 164
Westera, Matthijs, 152
Wham, Drew, 171
Wharton, Veronica, 48
White, Aaron Steven, 78
White, Michael, 92, 117, 169
Whyatt, Daniel, 82
Wickenberg Bolin, Ulrika, 153
Wiegand, Michael, 11, 81
Wiegreffe, Sarah, 95
Wieting, John, 122, 167
Wilkens, Rodrigo, 12, 18, 61
Willhelm, Chris, 49
Williams, Adina, 96, 137
Winter, Shuly, 11
Wiseman, Sam, 11
Wisniewski, Guillaume, 89, 105
Wojatzki, Michael, 167
Wong, Kam-Fai, 55
Wong, Yu-Shiang, 68
Wood-Doughty, Zach, 183, 184
Wu, Bowen, 142
Wu, Chuhan, 152, 154, 156, 162, 182
Wu, Chun-Kai, 84
Wu, Dapeng, 101
Wu, Fangzhao, 152, 154, 156, 162, 182
Wu, Hao, 86, 153, 155
Wu, Jiawei, 100
Wu, Lijun, 85
Wu, Sixing, 152, 154, 156, 162, 182
Wu, Xianchao, 141
Wu, Yuanbin, 154, 156, 162, 163
Xaobing, Zhou, 163
Xia, Jiangnan, 163
Xiao, Liqiang, 142
Xie, Ning, 18
Xie, Pengjun, 159
Xie, Stanley, 67
Xie, Yufei, 156
Xie, Ziang, 67
Xing, Eric, 68
Xiong, Wenhuan, 120
Xu, Bo, 102
Xu, Dongfang, 152
Xu, Frank, 179
Xu, Huimin, 154
Xu, Pauli, 97
Xu, Peng, 34, 154
Xu, Qiongkai, 156
Xu, Shaobin, 18, 179
Xu, Shuyao, 171
Xu, Wei, 46, 58
Xu, Weiran, 44
Xu, Yuan, 160
Xu, Zhen, 142
Xue, Nianwen, 61
Yadav, Shweta, 63
Yadav, Vikas, 166
Yaghoobzadeh, Yadollah, 179
Yamagishi, Hayahide, 106
Yan, Xiaohui, 57
Yan, Xifeng, 85, 120
Yan, Yonghong, 161
Yan, Zhao, 124
Yanaka, Hitomi, 84
Yaneva, Victoria, 172
Yang, Charles, 32
Yang, Dechuan, 156
Yang, Diyi, 16
Yang, Xiao, 171
Yang, Xutao, 163
Yang, Yang, 64
Yang, Yi, 84, 86

- Yang, Yinfei, 88, 127
Yang, Yiping, 154
Yang, Zhen, 102
Yang, Zhilin, 42
Yanagarber, Roman, 64
Yankov, Dragomir, 18
Yannakoudakis, Helen, 47, 169
Yasunaga, Michihiro, 91
Yates, Andrew, 175
Yatskar, Mark, 12, 42, 168
Yavuz, Semih, 85
Ye, Hai, 121
Yeh, Yi Ting, 44
Yi, Jinfeng, 100
Yifiru, Martha, 17
Yih, Wen-tau, 57, 62, 126, 143
Yimam, Seid Muhie, 169
Yin, Ruiqing, 180
Yin, Wenpeng, 12, 167
Yin, Zhenghang, 158
Yin, Zhongbo, 160
Yoon, Jong, 175
Yoon, Seunghyun, 125
Yoon, Su-Youn, 48
Yoshikawa, Masashi, 89
Yu, Dong, 163
Yu, Hong, 45
Yu, Hsiang-Fu, 80
Yu, Juntao, 185
Yu, Kai, 143
Yu, Mo, 100
Yu, Tao, 126
Yu, Yong, 34
Yu, Zhongzhi, 166
Yuan, Hang, 163
Yuan, Pei-Chieh, 144
Yuan, Zheng, 49, 171
Yuan, Zhigang, 152, 154, 156, 162, 182
Yuille, Alan, 55
Yuming, Ye, 160
Yvon, François, 89
- Zadeh, Amir, 143
Zafar, Lubna, 158
Zaiane, Osmar, 43
Zalmout, Nasser, 90
Zamani, Mohammadzaman, 175
Zampella, Casey, 175
Zampieri, Marcos, 169, 171
Zanzotto, Fabio Massimo, 156
Zaporojets, Klim, 175
Zaremoodi, Poorya, 104
Zargayouna, Haifa, 159
Zarrella, Guido, 163
Zayed, Omnia, 181
Zechner, Klaus, 48, 169
- Zehe, Albin, 160
Zeldes, Amir, 62, 185
Zeng, Xiaohui, 55
Zeng, Xingshan, 55
Zerva, Chrysoula, 176
Zesch, Torsten, 167, 172
Zettlemoyer, Luke, 36, 87, 122, 124, 137, 150
Zhan, Meilin, 139
Zhang, Alexander, 162
Zhang, Aston, 179
Zhang, Boliang, 12, 92, 170
Zhang, Ce, 159
Zhang, Chengzhi, 129
Zhang, Haisong, 58
Zhang, Haoran, 172
Zhang, Honglun, 142
Zhang, Jiajun, 11
Zhang, Jingyi, 103, 160
Zhang, Meng, 171
Zhang, Qiong, 86
Zhang, Rui, 126
Zhang, Shaodian, 34
Zhang, Sheng, 166
Zhang, Weinan, 34
Zhang, Xuejie, 154, 156, 158, 163
Zhang, Ye, 117
Zhang, Yi, 36
Zhang, Yingyi, 129
Zhang, You, 154
Zhang, Yuan, 54, 186
Zhang, Yue, 63, 125, 142
Zhang, Yunqi, 40
Zhang, Zheng, 180
Zhang, Zhengxin, 153
Zhang, Zhirui, 124, 130
Zhang, Zhousheng, 161
Zhang, Zilin, 126
Zhao, Hai, 161
Zhao, Huasha, 86
Zhao, Jieyu, 42
Zhao, Li, 85
Zhao, Meiqian, 163
Zhao, Wei, 159
Zhao, Xuemin, 161
Zhao, Yan, 163
Zheng, Huafei, 159
Zhou, Bowen, 100
Zhou, Deyu, 64, 131
Zhou, Jun, 127
Zhou, Liyuan, 156
Zhou, Ming, 124, 130
Zhou, Qimin, 153
Zhou, Xiaobing, 154, 163, 164
Zhou, Yunxiao, 162
Zhu, Huaiyu, 49
Zhu, Huiling, 126

- Zhu, Wei-Jing, 110
Zhu, Yi, 91
Zhunchen, Luo, 164
Zhuo, Hankz Hankui, 126
Ziai, Ramon, 41, 170
Ziebart, Brian, 19
Zimmermann, Roger, 63, 130, 143
Zinkov, Robert, 76
Zinsmeister, Heike, 185
Zirikly, Ayah, 174, 175
Ziser, Yftah, 99
Zitouni, Imed, 78
Zlatkova, Dimitrina, 157
Zopf, Markus, 130, 133
Zubiaga, Arkaitz, 11, 12
Zukerman, Ingrid, 99
Zündorf, Albert, 49
Zuylen, Madeleine van, 49, 128
Zweigenbaum, Pierre, 180

Google AI Residency Program

Google AI Residency Program

The Google AI Residency is a 12-month research training role designed to jumpstart your career in machine learning research. Residents will have the opportunity to work alongside distinguished scientists and engineers across various machine learning research teams within Google, such as **Google Perception**, **Natural Language Understanding**, **Google AI New York**, and **Google Brain team**. Residents will also have the opportunity collaborate with other Google and Alphabet teams such as **X**, **Google Cloud AI** and **Accelerated Sciences**. Check out g.co/airesidency for more information about our program, current resident profiles and examples of resident work.

Who should apply?

The ideal candidate should have a degree or equivalent practical experience in a STEM field such as Computer Science, Mathematics, or Statistics. Candidates should also present evidence of proficiency in programming (e.g. Python, C++) and in the prerequisite courses.

Most importantly, we are looking for people who are **motivated and want to learn to conduct machine learning research**. You may have research experience in another field (e.g. mathematics, physics, bioinformatics, etc.) and want to apply machine learning to this area, or you may have limited research experience but a strong desire to do more.

I'm interested! Can I apply now?

Applications for the 2019 AI Residency program will **open in Fall 2018**. Leave your details at g.co/airesidencysignup if you'd like to be notified when applications reopen!



GLOBAL CREATION AND INTERACTION PLATFORM

AI technology is at the heart of all of Bytedance's content platforms. We build intelligent machines that are capable of understanding and analyzing texts, images and videos using natural language processing and computer vision technology, then we use large-scale machine learning and deep learning algorithms to serve users with the content that they will find most interesting.



Toutiao



Tik Tok



Vigo Video



Xigua Video



musical.ly



News Republic



TopBuzz



BuzzVideo

lab-hr@bytedance.com



NAACL-HLT 2018

Data moves markets.

We build the systems
that understand it.

We're hiring

Machine Learning: bit.ly/MLatBBG

Artificial Intelligence: bit.ly/AI_Bloomberg

Email our recruiter Jen: jcarberry7@bloomberg.net
Follow us: @TechAtBloomberg

Engineering

Bloomberg

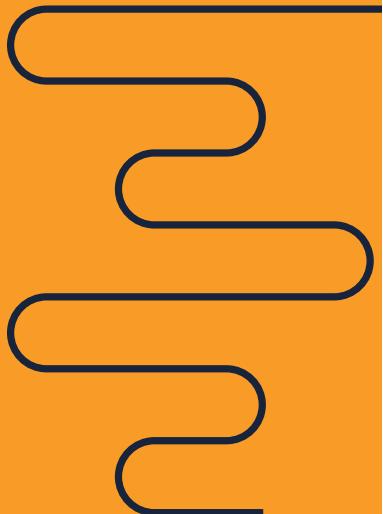
TechAtBloomberg.com

© 2018 Bloomberg Finance L.P. All rights reserved.

research



**It's only
impossible
until it
works**



**Learn more:
[AMAZON.JOBS/NAACL](https://amazon.jobs/naacl)**



WE PREDICT YOU'LL LIKE IT HERE

capitalonecareers.com



KPMG Ignite

Accelerating, automating, and augmenting decisions through Artificial Intelligence

KPMG Ignite includes methods, tools, approaches and resources that focus on improving the consistency, efficiency and time to make decisions and take action. In addition, we wrap this portfolio with highly-skilled resources who know how to apply relevant domain knowledge to leverage these capabilities and continue to innovate on emerging technologies.



Technology
Ecosystem



Technology
Workbench



Frameworks
and Methods



Teams



Research,
Innovation,
Development



Write the Future



- We push the boundaries of technology to improve how people communicate with each other.
- 10M+ users rely on Grammarly every day to make their writing clear, effective, and mistake-free.
- Our 150-person team in San Francisco, New York, and Kyiv is growing. Join us!

grammarly.com/jobs

We are Two Sigma

We find connections in all the world's data. These insights guide investment strategies that support retirements, fund research, advance education and benefit philanthropic initiatives. So, if harnessing the power of math, data and technology to help people is something that excites you, let's connect.

→ Visit us at www.twosigma.com/careers to learn more and apply.



 TWO SIGMA

NEW YORK | HOUSTON | LONDON | HONG KONG | TOKYO

@twosigmacareers
linkedin.com/company/two-sigma-investments
@twosigmacareers



BE A PART OF AN **EXCEPTIONAL PARTNERSHIP**

The newly established computer science department at Tulane University intends to be a leading program with an innovative curriculum that is known for education and research in the areas of Artificial Intelligence, Machine Learning, Natural Language Processing and other areas including multi-disciplinary research. Tulane's faculty have a close collaboration with IHMC, a leading research institute in AI, Robotics and Natural Language Understanding. **JOIN US!**



<https://www2.tulane.edu/sse/cs/>



The Platform for Conversational AI



Machine learning is the lifeblood of virtual personal assistants. Today, only the tech giants have the AI arsenal to build good voice-based experiences. At PolyAI, we are developing a machine learning platform which will make conversational app design as simple as building a website.

If you would like to hear more, come to our tutorial on *Deep Learning for Conversational AI*, and visit our NAACL booth.

Our team has been publishing at *ACL conferences for many years. With PolyAI, we are taking our research prototypes to a whole new level. We are currently expanding our teams in London and Singapore. If you're interested in *dialogue, language understanding, generation or reinforcement learning*, **get in touch**.



Deep Learning for NLP
Robust language understanding and rich response generation



Multilingual
Our models are polyglots, scaling across languages at near-zero cost



Task Agnostic
Our platform is the same for any application, improving as it scales across different tasks and domains



High-Quality Data
On-demand data collection with a proprietary crowdsourcing platform



IBM Research is hiring

We live in a moment of remarkable change and opportunity. Data and technology are transforming industries and societies, ushering in a new era of Artificial Intelligence. IBM Research is a leader in this worldwide transformation, building on a long history of innovation.

For more than seven decades, IBM Research has defined the future of technology. Our scientists, among them six Nobel Laureates and six Turing Award winners, have produced ten U.S. National Medals of Technology and five U.S. National Medals of Science. Along the way we helped put a man on the moon, defeated Kasparov at chess, and built a Jeopardy! champion named Watson.

At IBM, you can achieve what others think is impossible. And in doing so, you'll play a significant role in shaping the future. Join us.

Discover what you can do at IBM Research
ibm.com/jobs

© Copyright IBM Corporation 2018. IBM, the IBM logo and ibm.com are trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. See current list at ibm.com/trademark. Other product and service names might be trademarks of IBM or other companies.

#1 Business Software

Complete. Modern. Integrated.

ORACLE®

oracle.com/about

Copyright © 2018, Oracle and/or its affiliates. All rights reserved.

Inventing a more human conversation with technology

To learn more about Nuance's R&D team of talented scientists, linguists and engineers pioneering human-machine intelligence and communication, visit research.nuance.com



Facebook Fellowship Program & Emerging Scholar Awards

Facebook Fellowships and Emerging Scholar Awards are designed to encourage and support promising doctoral students engaged in innovative and relevant research.

Learn more at
research.fb.com/programs/

Applications open
September 4, 2018

facebook research

An advertisement for the Information Sciences Institute (ISI). It features a large, stylized graphic of two hands reaching towards each other, one blue and one red, set against a dark background with a hexagonal grid pattern. The text "AI @ ISI" is prominently displayed in yellow. Below the hands, a vertical list of research areas includes: Natural Language Processing, Information Integration, Discovery Informatics, Dynamic Networks, Machine Learning, Discourse Analysis, Hybrid Reasoning, Sentence Meaning, Decipherment, Neuroscience, Text Generation, Semantic Web, Machine Translation, Multilingual Text Mining, Scientific Workflows, Trust, Crowdsourcing, Provenance, Knowledge Capture, Econometrics, Big Data, Linked Data, Geospatial Computing, Intelligent Interfaces, Assistive Robotics, Human Behavior, Bioinformatics, Knowledge Engineering, Human Cognition, Information Theory, Link Discovery, Quantum Theory, Social Media, Statistical Inference, and Information Propagation.

JOIN US!

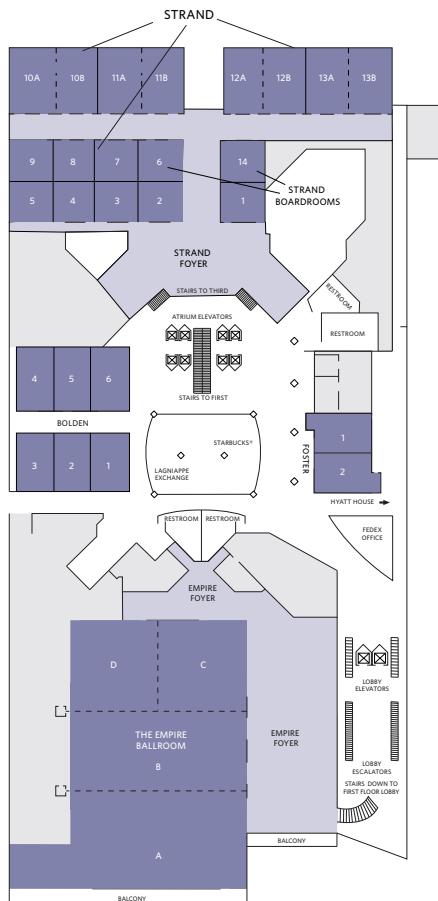
Research Scientist - Postdoc - PhD Student

Visiting / Sabbatical Researcher

Senior Research Fellow - Research Programmer - Summer Intern

Information Sciences Institute

An advertisement for iMerit. The top half features the iMerit logo with the tagline "human-empowered computing". Below the logo, the text "Make your NLP data work" is displayed. The bottom half shows a stylized profile of a human head containing a network of connections. Inside the head, there is a smartphone displaying the word "Pastry" and the text "NOLA 78°F". Various icons representing different NLP applications like weather, time, and location are scattered around the head and phone.



We gratefully acknowledge the sponsors of NAACL HLT 2018!

Diamond Sponsors



Platinum Sponsors



Gold Sponsors



Silver Sponsors



Bronze Sponsors



Sponsors of the Student Research Workshop

