

ICON-2018

**15th International
Conference on Natural
Language Processing**

Proceedings of the Conference

15-18 December 2018
Punjabi University, Patiala, India

© 2018 NLP Association of India (NLPAI)

Preface

Natural Language Processing (NLP) research has evolved from the era of punch cards and batch processing, in which the analysis of a sentence could take several minutes, to the era of Google and the likes of it, in which millions of webpages can be processed in matter of milliseconds. NLP enables computers to perform a wide range of natural language related tasks at all levels, ranging from parsing and opinion mining, to machine translation and speech recognition. Research in Natural Language Processing (NLP) has taken a noticeable leap in the recent years. Tremendous growth of information on the web and its easy access has stimulated large interest in the field. With the ongoing growth of the World Wide Web and social media, there is a drastic increase in online data. As the amount of data increases the mechanisms to process these unstructured data and to extract meaningful information from it becomes more challenging. These challenges and difficulties can be overcome with the advanced NLP techniques. Deep learning architectures and algorithms are making impressive advances in NLP. Recent NLP research is now increasingly focusing on the use of new deep learning methods.

Indic language processing presents formidable challenges to achieving multilingualism and multiculturalism in the Indian subcontinent. One of the first and most obvious challenges is the multitude and diversity of languages: India is a land of diverse culture with around 33 major languages and 1,652 dialects from half-a-dozen different language groups. Indic Language Processing involves developing software in Indic Scripts/ languages, Input methods, Localization of computer applications, Web development, Internationalized Domain Name (IDN), OCR, Spell-checkers, Speech applications etc. Research on Indian language technology has thrived in the past few years, with ICON (International Conference on Natural Language Processing), being the premier conference in this field.

This volume contains papers selected for presentation in technical sessions of 15th International Conference on Natural Language Processing (ICON-2018) and short communications selected for poster presentation. We are thankful to our excellent team of reviewers from all over the globe who deserve full credit for the hard work put in for reviewing the high quality submissions with rich technical content. From 152 submissions, 27 papers have been finally selected, 14 for oral and 13 for poster presentation, representing a variety of new and interesting developments, covering a wide spectrum of NLP areas and core linguistics.

We are deeply grateful to Dr. Pushpak Bhattacharyya, Director and Professor of Computer Science and Engineering, IIT Patna, India and Mr. Sanjeev Gupta, Flipkart, Bangalore, for delivering the keynote lectures at ICON-2018. We would also like to thank the members of the Advisory Committee and Programme Committee for their support and co-operation in making ICON 2018 a success.

We thank Prof. Sanjay Dwivedi and Dr. Sarika Jain, Chair, Student Paper Competition and Dr. M D Kulkarni and Dr. Rajeev R R, Chair, NLP Tools Contest for taking the responsibilities of the events.

We are also grateful to Dr. Anil Kumar Singh for devoting his precious time to shape this volume in its present form. His timely help enable us to release this volume in time.

We convey our thanks to Balwant Singh, Dharwinder Singh, Rakesh Dawra from Punjabi University Patiala and P V S Ram Babu from International Institute of Information Technology (IIIT), Hyderabad for their dedicated efforts in successfully handling the ICON Secretariat. We heartily express our

gratitude to the dedicated team of volunteers at Punjabi University Patiala and the entire staff of Research Centre for Technical Development of Punjabi Language Literature and Culture, Punjabi University Patiala for working tirelessly day and night for the success of the conference.

We also thank all those who came forward to help us in this task. We apologize if we have missed some names.

Finally, we thank all the researchers who responded to our call for papers and all the participants of ICON-2018, without whose overwhelming response the conference would not have been a success.

December 2018
Patiala

Gurpreet Singh Lehal
Dipti Misra Sharma
Rajeev Sangal

Conference General Chair:

Rajeev Sangal, IIT (BHU), Varanasi, India

Programme Committee:

Kalika Bali, Microsoft Research, India
Srinivas Bangalore, Interactions LLC, AT&T Research, USA
Rajesh Bhatt, University of Massachusetts, USA
Pushpak Bhattacharyya, IIT Patna, India
Monojit Choudhury, Microsoft Research, India
Harald Hammarström, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
Mohammed Hasanuzzaman, Université de Caen, Normandie, France
Jyoti Pawar, DCST, Goa University, India
L Ramamoorthy, CIIL Mysore, India
Owen Rambow, University of Columbia, USA
Elizabeth Sherly, IIITM-K, Trivandrum, India
Keh-Yih Su, Institute of Information Science, Academia Sinica, Taiwan
Vasudeva Varma, IIIT Hyderabad, India

NLP Software Chairs:

M. D. Kulkarni, CDAC Pune, India (Chair)
Rajeev R.R., ICFOSS Thiruvananthapuram, India (Co-Chair)

Student Paper Competition Chair:

Sanjay Dwivedi, BBAU (Central University Lucknow), India (Chair)
Sarika Jain, NIT Kurukshetra, India (Co-Chair)

Organizing Committee:

Vishal Goyal, Punjabi University Patiala, India
Gurpreet Singh Joshan, Punjabi University Patiala, India
Tejinder Singh Saini, Punjabi University Patiala, India
Ankur Rana, Punjabi University Patiala, India

Sponsors

Microsoft®
Research

Flipkart



विज्ञान एवं प्रौद्योगिकी विभाग
DEPARTMENT OF
SCIENCE & TECHNOLOGY

Referees

We gratefully acknowledge the excellent quality of refereeing we received from the reviewers. We thank them all for being precise and fair in their assessment and for reviewing the papers in time.

Adithya Pratapa
Ajit Kumar
Amol Bole
Anchal Rani
Ankur Rana
Anoop Kunchukuttan
Anshul Bawa
Anupam Jamatia
Anupam Mondal
Anushiya G
ANZI A S
Arjun Akula
Arun Baby
Aswin Shanmugam
Atul Ojha
Avijit Thawani
Bankim Patel
Basant Agarwal
Bhubvaneshwari Melinamath
Braja Patra
C Chandra Shekhar
Deepali Singla
Delia Hernández Farias
Devi G
Dhanya K
Dinesh Kumar
Dr. B. Pawar

Dr.Preeti Dubey
Dwijen Rudrapal
Golda Rajan
Gurpreet Josan
Gurpreet Lehal
K Mishra
Kamal Garg
Kamal Sarkar
Kevin Patel
Kunal Chakma
Latha Nair
Lovely Sharma
Mamoon Rashid
Mayank Singh
Niloofar Safi
Nirmal Surange
Parminder Singh
Parteek Bhatia
Partha Pakray
Pranav Dhakras
Pranav Goel
Prateek Agrawal
Rajeev Puri
Rajiv Srivastava
Ravinder Kumar
Rudramurthy V
Sachin Pawar

Sachin Ruikar
Sanjeev Sharma
Santanu Pal
Shanky Goel
Shashank Gupta
Shilpa Desai
Shubhnandan Jamwal
Sivanand Achanta
Sopan Kolte
Sourabh Kumar
Sourav Mandal
Subba Oota
Sudipta Kar
Tejinder Singh Saini
Thoudam Singh
Umamaheswari E
Umrinder Singh
V Singh
Vasudeva Varma
Venkata Viraraghavan
Vijay Ram
Vijayalakshmi P
Vimal Soni
Vishal Goyal
Vishal Gupta
Yan Shao

Table of Contents

<i>Abstractive Summarization Using Attentive Neural Techniques</i> Jacob Krantz and Jugal Kalita	1
<i>Challenges and Issues in Developing an Annotated Corpus and HMM POS Tagger for Khasi</i> Medari Janai Tham	10
<i>Infant Crying Cause Recognition using Conventional and Deep Learning based Approaches</i> Shivam Sharma, P. Viswanath and Vinay Kumar Mittal	20
<i>Automatic Pause Boundary and Pause Duration Detection for Text-to-Speech Synthesis Systems in Indian Languages</i> Atish Shankar Ghone, Rachana Nerpagar, Pranaw Kumar, BiraChandra Singh, Prakash B. Pimpale and Sasikumar M.	28
<i>Summarization of Table Citations from Text</i> Monalisa Dey, Salma Mandi and Dipankar Das	35
<i>Khasi Shallow Parser</i> Medari Janai Tham	43
<i>Resolving Actor Coreferences in Hindi Narrative Text</i> Nitin Ramrakhiyani, Swapnil Hingmire, Sachin Pawar, Sangameshwar Patil, Girish K. Palshikar, Pushpak Bhattacharyya and Vasudeva Verma	50
<i>Deep Learning methods for Semantic Role Labeling in Indian Languages</i> Aishwary Gupta, Akshay Pawale and Manish Shrivastava	59
<i>Machine Learning Approaches for Amharic Parts-of-speech Tagging</i> Ibrahim Gashaw and H. L. Shashirekha	69
<i>English to Bodo Statistical Machine Translation System Using Multi-domain Parallel Corpora</i> Saiful Islam	75
<i>POS Tagging and Named Entity Recognition on Handwritten Documents</i> Vijay Rowtula and Praveen Krishnan	82
<i>A New Chat Solution for Shared Services using Natural Language Processing Models</i> M. Saravanan, Satheesh K. Perepu and Sudipta Bose	87
<i>Does Curriculum Learning help Deep Learning for Natural Language Generation?</i> Sandhya Singh, Kevin Patel, Pushpak Bhattacharya, Krishnanjan Bhattacharjee, Hemant Darbari and Seema Verma	92
<i>WupLeBleu: The Word-net Based Evaluation Metric for Machine Translation</i> Debajyoty Banik, Asif Ekbal and Pushpak Bhattacharyya	99
<i>"Is This A Joke?": A Large Humor Classification Dataset</i> Faraz Faruqi and Manish Shrivastava	104

<i>Fuzzy Evolutionary Self-Rule generation and Text Summarization</i> Pradeepika Verma and Hari Om	110
<i>A Content-based Recommendation System for Medical Concepts: Disease and Symptom</i> Anupam Mondal, Dipankar Das and Sivaji Bandyopadhyay	115
<i>A Deep Learning Model for Event Extraction and Classification in Hindi for Disaster Domain</i> Zishan Ahmad, Sahoo Sovan Kumar, Asif Ekbal and Pushpak Bhattacharyya	122
<i>Exploring the importance of context and embeddings in neural NER models for task-oriented dialogue systems</i> Pratik Jayarao, Chirag Jain and Aman Srivastava	132
<i>Improving Computer Generated Dialog with Auxiliary Loss Functions and Custom Evaluation Metrics</i> Thomas Conley , Jack St. Clair and Jugal Kalita	138
<i>Analyzing Autism Speech of Children in English Vowels Regions by Analysis of Changes in Production Features</i> Abhijit Mohanta and Vinay Kumar Mittal	145
<i>Hate Speech Detection from Code-mixed Hindi-English Tweets Using Deep Learning Models</i> Satyajit Kamble and Aditya Joshi	150
<i>Semi-Supervised Confidence Network aided Gated Attention based Recurrent Neural Network for Click-bait Detection</i> Amrith Rajagopal Setlur	156
<i>User Perception of Code-Switching Dialog Systems</i> Anshul Bawa, Monojit Choudhury and Kalika Bali	166
<i>SMT vs NMT: A Comparison over Hindi and Bengali Simple Sentences</i> Sainik Kumar Mahata, Soumil Mandal, Dipankar Das and Sivaji Bandyopadhyay	175
<i>Helping each Other: A Framework for Customer-to-Customer Suggestion Mining using a Semi-supervised Deep Neural Network</i> Hitesh Golchha, Deepak Gupta, Asif Ekbal and Pushpak Bhattacharyya	183
<i>Towards Predicting Age of Acquisition of Words Using a Dictionary Network</i> Ditty Mathew, Girish Raguvir Jeyakumar, Rahul Kejriwal and Sutanu Chakraborti	193

Conference Program

Sunday, December 16, 2018

+ 8:00-9:30 Registration

+ 9:30-11:00 Inaugural Ceremony

+ 9:30-11:00 Keynote Lecture 1 by Dr. Pushpak Bhattacharya

+ 11:00-11:30 Tea Break

+ 11:30-13:00 Technical Session I : Speech and Summarization

Abstractive Summarization Using Attentive Neural Techniques

Jacob Krantz and Jugal Kalita

Challenges and Issues in Developing an Annotated Corpus and HMM POS Tagger for Khasi

Medari Janai Tham

Infant Crying Cause Recognition using Conventional and Deep Learning based Approaches

Shivam Sharma, P. Viswanath and Vinay Kumar Mittal

Automatic Pause Boundary and Pause Duration Detection for Text-to-Speech Synthesis Systems in Indian Languages

Atish Shankar Ghone, Rachana Nerpagar, Pranaw Kumar, BiraChandra Singh, Prakash B. Pimpale and Sasikumar M.

+ 13:00-14:00 Lunch

Sunday, December 16, 2018 (continued)

+ 14:00-15:30 Technical Session II : NLP Tools

Summarization of Table Citations from Text

Monalisa Dey, Salma Mandi and Dipankar Das

Khasi Shallow Parser

Medari Janai Tham

Resolving Actor Coreferences in Hindi Narrative Text

Nitin Ramrakhiani, Swapnil Hingmire, Sachin Pawar, Sangameshwar Patil, Girish K. Palshikar, Pushpak Bhattacharyya and Vasudeva Verma

Deep Learning methods for Semantic Role Labeling in Indian Languages

Aishwary Gupta, Akshay Pawale and Manish Shrivastava

+ 15:30-15:45 Tea Break

+ 15:45-17:45 Poster Session

Machine Learning Approaches for Amharic Parts-of-speech Tagging

Ibrahim Gashaw and H. L. Shashirekha

English to Bodo Statistical Machine Translation System Using Multi-domain Parallel Corpora

Saiful Islam

POS Tagging and Named Entity Recognition on Handwritten Documents

Vijay Rowtula and Praveen Krishnan

A New Chat Solution for Shared Services using Natural Language Processing Models

M. Saravanan, Satheesh K. Perepu and Sudipta Bose

Does Curriculum Learning help Deep Learning for Natural Language Generation?

Sandhya Singh, Kevin Patel, Pushpak Bhattacharya, Krishnanjan Bhattacharjee, Hemant Darbari and Seema Verma

WupLeBleu: The Word-net Based Evaluation Metric for Machine Translation

Debajyoty Banik, Asif Ekbal and Pushpak Bhattacharyya

Sunday, December 16, 2018 (continued)

"Is This A Joke?": A Large Humor Classification Dataset

Faraz Faruqi and Manish Shrivastava

Fuzzy Evolutionary Self-Rule generation and Text Summarization

Pradeepika Verma and Hari Om

A Content-based Recommendation System for Medical Concepts: Disease and Symptom

Anupam Mondal, Dipankar Das and Sivaji Bandyopadhyay

A Deep Learning Model for Event Extraction and Classification in Hindi for Disaster Domain

Zishan Ahmad, Sahoo Sovan Kumar, Asif Ekbal and Pushpak Bhattacharyya

Exploring the importance of context and embeddings in neural NER models for task-oriented dialogue systems

Pratik Jayarao, Chirag Jain and Aman Srivastava

Improving Computer Generated Dialog with Auxiliary Loss Functions and Custom Evaluation Metrics

Thomas Conley , Jack St. Clair and Jugal Kalita

Analyzing Autism Speech of Children in English Vowels Regions by Analysis of Changes in Production Features

Abhijit Mohanta and Vinay Kumar Mittal

+ 18:00-19:00 NLP AI Meeting

+ 19:00-20:30 Cultural Programme

+ 20:30-Onwards Dinner

Monday, December 17, 2018

+ 9:30-10:15 Industry Session Talk by Sanjeev Gupta

+ 10:15-10:30 Tea Break

+ 10:30-12:00 Technical Session III: Social Media

Hate Speech Detection from Code-mixed Hindi-English Tweets Using Deep Learning Models

Satyajit Kamble and Aditya Joshi

Semi-Supervised Confidence Network aided Gated Attention based Recurrent Neural Network for Clickbait Detection

Amrith Rajagopal Setlur

User Perception of Code-Switching Dialog Systems

Anshul Bawa, Monojit Choudhury and Kalika Bali

+ 12:00-13:00 NLP Software Contest Session-I

+ 13:00-14:00 Lunch

+ 14:00-15:45 NLP Software Contest Session-II

+ 15:45-16:00 Tea

+ 16:00-18:00 Technical Session IV : Recent Trends

SMT vs NMT: A Comparison over Hindi and Bengali Simple Sentences

Sainik Kumar Mahata, Soumil Mandal, Dipankar Das and Sivaji Bandyopadhyay

Helping each Other: A Framework for Customer-to-Customer Suggestion Mining using a Semi-supervised Deep Neural Network

Hitesh Golchha, Deepak Gupta, Asif Ekbal and Pushpak Bhattacharyya

Towards Predicting Age of Acquisition of Words Using a Dictionary Network

Ditty Mathew, Girish Raguvir Jeyakumar, Rahul Kejriwal and Sutanu Chakraborti

Monday, December 17, 2018 (continued)

+ 18:00-18:30 Valedictory Function

