COLING 2012

24th International Conference on Computational Linguistics

Proceedings of the
First International Workshop on Optimization Techniques for Human Language Technology

Workshop chairs:
Pushpak Bhattacharyya, Asif Ekbal, Sriparna Saha,
Mark Johnson, Diego Molla-Aliod and Mark Dras

9 December 2012

Diamond sponsors

Tata Consultancy Services
Linguistic Data Consortium for Indian Languages (LDC-IL)

Gold Sponsors

Microsoft Research
Beijing Baidu Netcon Science Technology Co. Ltd.

Silver sponsors

IBM, India Private Limited
Crimson Interactive Pvt. Ltd.
Yahoo
Easy Transcription & Software Pvt. Ltd.

Preface

In decision science, optimization is quite an obvious and important tool. Depending on the number of objectives, the optimization technique can be single or multiobjective. We encounter numerous real life scenarios where multiple objectives need to be satisfied in the course of optimization. Finding a single solution in such cases is very difficult, if not impossible. In such problems, referred to as multiobjective optimization problems (MOOPs), it may also happen that optimizing one objective leads to some unacceptably low value of the other objective(s). Evolutionary algorithms and simulated annealing, from the family of meta-heuristic search and optimization techniques, have shown promise in solving complex single as well as multiobjective optimization problems in a wide variety of domains.

Language technology and/or Natural language processing (NLP) is an interdisciplinary field of computer science and linguistics concerned with the interactions between computers and human (natural) languages. It is a branch of artificial intelligence. In theory, NLP is a very attractive method of human-computer interaction. Natural language understanding is sometimes referred to as an AI-complete problem because it seems to require extensive knowledge about the outside world and the ability to manipulate it. Modern NLP algorithms are grounded in machine learning, especially statistical machine learning. Research into modern statistical NLP algorithms requires an understanding of a number of disparate fields, including linguistics, computer science, and statistics. Major tasks in NLP include Automatic summarization, Coreference resolution, Named Entity Recognition, Machine translation, Machine transliteration, Natural language generation, Natural language understanding, Morphological segmentation, Part-of-Speech tagging, Question answering, Sentiment analysis, Speech segmentation, Word sense disambiguation, Information retrieval etc.

In each of the above mentioned tasks, there are various metrics that we often need to optimize to get the reasonable performance. Many evaluation metrics have been proposed for solving different problems of NLP. For example, in Information retrieval, it is often necessary to optimize the recall and precision parameters. In automatic summarization, it is desired to optimize different objective functions like similarity to user query, ROUGE metric, important sentence score, difference in length between the scored sentence and the desired sentence and many others. Other examples of optimization in NLP include parsing, machine translation, and computational models of language acquisition.

This volume contains papers accepted for presentation at the First International Workshop on Optimization Techniques for Human Language Technology. The event took place on December 9, 2012, in Mumbai, India, as a workshop in COLING 2012, the 24th International Conference on Computational Linguistics. The workshop was a starting platform to explore the possibilities of interdisciplinary research that will focus on developing optimisation based methods within the context of human language technology.

Eight papers were accepted for presentation, based on the careful reviews of a panel of international experts in various areas related to the workshop goals. Our sincere thanks to all the reviewers for their thoughtful reviews.

We would like to thank Prof. Aravind K. Joshi, University of Pennsylvania for his invited speech on "Complexity of Parse representations, Parsing complexity, Side Information: Relevance to Optimization?"

Asif Ekbal, Pushpak Bhattacharyya, Sriparna Saha,
Mark Johnson, Diego Molla, Mark Dras.
December 2012.

Organizers:

Asif Ekbal (Indian Institute of Technology Patna, Bihar, India) (Chair)
Pushpak Bhattacharyya (Indian Institute of Technology Bombay, India)
Sriparna Saha (Indian Institute of Technology Patna, India)
Mark Johnson (Department of Computing, Macquarie University, Australia)
Diego Molla (Macquarie University, Australia)
Mark Dras (Macquarie University, Australia)

Program Committee:

Ramiz M. Aliguliyev (Azerbaijan National Academy of Sciences, Azerbaijan)
Timothy Baldwin (University of Melbourne, Australia)
Sivaji Bandyopadhyay (Jadavpur University, India)
Malay Bhattacharyya (Kalyani University, India)
Pushpak Bhattacharyya (Indian Institute of Technology Bombay, India)
Benjamin Boerschingen (Macquarie University, Australia)
Niladri Chatterjee (IIT Delhi)
Monojit Choudhury (Microsoft Research India)
Walter Daelemans (University of Antwerp)
Gaël Harry Dias (University of Caen Basse-Normadie, France)
Soumyajit Dey (Indian Institute of Technology Patna, India)
Patrick Saint-Dizier (Institut de Recherches en Informatique de Toulouse)
Mark Dras (Macquarie University, Australia)
Lan Du (Macquarie University, Australia)
Asif Ekbal (Indian Institute of Technology Patna, Bihar, India) (Chair)
Alexander Gelbukh (National Polytechnic Institute (IPN), Mexico)
Veronique Hoste (University College Ghent)
Jagadeesh Jagarlamudi (University of Maryland College Park, USA)
Mark Johnson (Department of Computing, Macquarie University, Australia)
Nitin Indurkya (University of New South Wales, Australia)
Zornitsa Kozareva (Information Sciences Institute / University of Southern California, USA)
A Kumaran (Microsoft Research India)
Pabitra Mitra (Indian Institute of Technology Kharagpur, India)
Diego Molla (Macquarie University, Australia)
Samrat Mondal (Indian Institute of Technology Patna, India)
Anirban Mukhopadhyay (Kalyani University, India)
Massimo Poesio (University of Trento, Italy)
Sriparna Saha (Indian Institute of Technology Patna, India)
Ashok Singh Sairam (Indian Institute of Technology Patna)
Soumi Sengupta (Indian Statistical Institute Kolkata, India)
Jyoti Prakash Singh (National Institute of Technology Patna, India)
Olga Uryupina (University of Trento, Italy)
Sriram Venkatapathy (Xerox Research Centre Europe)
Jose Luis Vicedo (University of Alicante, Spain)

Invited Speaker:

Aravind K. Joshi (University of Pennsylvania, USA)

Table of Contents

First International Workshop on Optimization Techniques for Human Language Technology

Program

Sunday, 9 December 2012

09:45	Start
10:00–11:00	**Invited Talk:** *Complexity of Parse representations, Parsing complexity, Side Information: Relevance to Optimization?* Aravind K. Joshi, University of Pennsylvania
	Session 1
11:00–11:30	*BioPOS: Biologically Inspired Algorithms for POS Tagging* Ana Paula Silva, Arlindo Silva and Irene Rodrigues
11:30–12:00	Tea break
	Session 2
12:00–12:30	*Optimization for Efficient Determination of Chunk in Automatic Evaluation for Machine Translation* Hiroshi Echizen'ya, Kenji Araki and Eduard Hovy
12:30–13:00	*Optimizing Transliteration for Hindi/Marathi to English Using only Two Weights* Manikrao Dhore, Shantanu Dixit and Ruchi Dhore
13:00–13:30	*Selection of Discriminative Features for Translation Texts* Kuo-Ming Tang, Chien-Kang Huang and Chia-Ming Lee
13:30–14:30	Lunch
	Session 3
14:30–15:00	*Semi-supervised Learning of Naive Bayes Classifier with feature constraints* Nagesh Bhattu Sristy and D.V.L.N Somayajulu
15:00–15:30	*Optimization and Sampling for NLP from a Unified Viewpoint* Marc Dymetman, Guillaume Bouchard and Simon Carter
15:30–16:00	*Iterative Chinese Bi-gram Term Extraction Using Machine-learning Classification Approach* Chia-Ming Lee, Chien-Kang Huang and Kuo-Ming Tang
16:00–16:30	*Parameter estimation under uncertainty with Simulated Annealing applied to an ant colony based probabilistic WSD algorithm* Andon Tchechmedjiev, Jérôme Goulian, Didier Schwab and Gilles Sérasset
16:30	Workshop end