

Tutorial: ‘*Computational Sarcasm*’

Pushpak Bhattacharyya¹ Aditya Joshi^{1,2,3}

¹IIT Bombay, India

²Monash University, Australia

³IITB-Monash Research Academy, India

{pb, adityaj}@cse.iitb.ac.in

1 Description

Sarcasm is a form of verbal irony that is intended to express contempt or ridicule. Motivated by challenges posed by sarcastic text to sentiment analysis, computational approaches to sarcasm have witnessed a growing interest at NLP forums in the past decade. Computational sarcasm refers to automatic approaches pertaining to sarcasm. The tutorial will provide a bird’s-eye view of the research in computational sarcasm for text, while focusing on significant milestones.

The tutorial begins with linguistic theories of sarcasm, with a focus on incongruity: a useful notion that underlies sarcasm and other forms of figurative language. Since the most significant work in computational sarcasm is sarcasm detection: predicting whether a given piece of text is sarcastic or not, sarcasm detection forms the focus hereafter. We begin our discussion on sarcasm detection with datasets, touching on strategies, challenges and nature of datasets. Then, we describe algorithms for sarcasm detection: rule-based (where a specific evidence of sarcasm is utilised as a rule), statistical classifier-based (where features are designed for a statistical classifier), a topic model-based technique, and deep learning-based algorithms for sarcasm detection. In case of each of these algorithms, we refer to our work on sarcasm detection and share our learnings. Since information beyond the text to be classified, contextual information is useful for sarcasm detection, we then describe approaches that use such information through conversational context or author-specific context.

We then follow it by novel areas in computational sarcasm such as sarcasm generation, sarcasm v/s irony classification, etc. We then summarise the tutorial and describe future directions based on errors reported in past work. The tutorial will end with a demonstration of our work on sarcasm detection.

This tutorial will be of interest to researchers investigating computational sarcasm and related areas such as computational humour, figurative language understanding, emotion and sentiment sen-

timent analysis, etc. The tutorial is motivated by our continually evolving survey paper of sarcasm detection, that is available on arXiv at: Joshi, Aditya, Pushpak Bhattacharyya, and Mark James Carman. ”Automatic Sarcasm Detection: A Survey.” arXiv preprint arXiv:1602.03426 (2016).

2 Tutorial Outline

This outline has been planned for 180 minutes. (95 minutes + 85 minutes = 180 minutes)

1. Introduction (15 minutes)
 - Sentiment analysis & Computational Sarcasm (3 minutes)
 - Motivation: Application to chatbots and sentiment analysis (3 minutes)
 - Challenges of computational sarcasm (5 minutes)
 - Scope of the tutorial (4 minutes)
2. Linguistic Foundations of Sarcasm (15 minutes)
 - Linguistic formulations of sarcasm (5 minutes)
 - Relationship between sarcasm, irony, deception, humor and metaphor (5 minutes)
 - Incongruity as a key indicator of sarcasm (5 minutes)
3. Datasets for Sarcasm detection (15 minutes)
 - Nature of datasets (5 minutes)
 - Strategies of annotation (5 minutes)
 - Challenges in obtaining labeled datasets (hashtag-based supervision, impact of non-native annotators, etc.) (5 minutes)
4. Algorithms for Sarcasm detection - 1 (50 minutes)
 -
 - Rule-based algorithms (15 minutes)
 - Past rule-based algorithms
 - Our work on rule-based algorithms

- Statistical classifier-based algorithms (35 minutes)
 - Features
 - Learning algorithms
 - Our work on classifier-based algorithms
5. -Coffee Break-
 6. Algorithms for Sarcasm detection - 2 (30 minutes)
 - Topic modeling-based algorithms (10 minutes)
 - Deep learning-based algorithms for sarcasm detection (20 minutes)
 - Past deep learning-based work
 - Our work on deep learning-based algorithms
 7. Contextual Information for Sarcasm Detection (15 minutes)
 - Conversational context (10 minutes)
 - Author’s historical context (5 minutes)
 8. Beyond Sarcasm Detection (20 minutes)
 - Sarcasm versus irony detection (6 minutes)
 - Sarcasm generation (6 minutes)
 - Sarcasm Target Identification (8 minutes)
 9. Summary (5 minutes)
 10. Future Directions (10 minutes)
 - Peculiar forms of sarcasm (pointers to sarcasm detection based on error analysis in past work.) (5 minutes)
 - Contextual information for sarcasm (5 minutes)
 11. Demonstration (5 minutes)

3 Speaker Profiles

1. Dr. Pushpak Bhattacharyya

Indian Institute of Technology Bombay, Mumbai, India & Indian Institute of Technology Patna, Patna, India.

pb@cse.iitb.ac.in, pushpakbh@gmail.com

www.cse.iitb.ac.in/~pb/

Prof. Pushpak Bhattacharyya is the current President of ACL (2016-17). He is the Director of IIT Patna and Vijay and Sita Vashee Chair Professor in IIT Bombay, Computer Science and Engineering Department. He was educated in IIT Kharagpur (B.Tech), IIT Kanpur (M.Tech) and IIT Bombay (PhD). He has been visiting scholar and faculty in MIT, Stanford, UT Houston and

University Joseph Fouriere (France). Prof. Bhattacharyya’s research areas are Natural Language Processing, Machine Learning and AI. He has guided more than 250 students (PhD, masters and Bachelors), has published more than 250 research papers and led government and industry projects of international and national importance. A significant contribution of his is Multilingual Lexical Knowledge Bases and Projection. Author of the text book ‘Machine Translation’ Prof. Bhattacharyya is loved by his students for his inspiring teaching and mentorship. He is a Fellow of National Academy of Engineering and recipient of Patwardhan Award of IIT Bombay and VNMM award of IIT Roorkey- both for technology development, and faculty grants of IBM, Microsoft, Yahoo and United Nations.

2. Aditya Joshi

IITB-Monash Research Academy, Mumbai, India. (IIT Bombay, India, & Monash University, Australia)

adityaj@cse.iitb.ac.in, aditya.m.joshi@gmail.com

<http://www.cse.iitb.ac.in/~adityaj/>

Aditya Joshi is a PhD student at IITB-Monash Research Academy, a joint PhD programme between Indian Institute of Technology Bombay, India and Monash University, Australia, since January 2013. His PhD advisors are Pushpak Bhattacharyya (IITB) and Mark Carman (Monash). His primary research focus is computational sarcasm where he has explored different ways in which incongruity can be captured in order to detect and generate sarcasm. In addition, he has worked on innovative applications of NLP such as sentiment analysis for Indian languages, drunk-texting prediction, news headline translation, political issue extraction, etc.