

# Cognitive Systems Monographs

Volume 37

## Series editors

Rüdiger Dillmann, University of Karlsruhe, Karlsruhe, Germany  
e-mail: ruediger.dillmann@kit.edu

Yoshihiko Nakamura, Tokyo University, Tokyo, Japan  
e-mail: nakamura@ynl.t.u-tokyo.ac.jp

Stefan Schaal, University of Southern California, Los Angeles, USA  
e-mail: sschaal@usc.edu

David Vernon, University of Skövde, Skövde, Sweden  
e-mail: david@vernon.eu

The Cognitive Systems Monographs (COSMOS) publish new developments and advances in the fields of cognitive systems research, rapidly and informally but with a high quality. The intent is to bridge cognitive brain science and biology with engineering disciplines. It covers all the technical contents, applications, and multidisciplinary aspects of cognitive systems, such as Bionics, System Analysis, System Modelling, System Design, Human Motion, Understanding, Human Activity Understanding, Man-Machine Interaction, Smart and Cognitive Environments, Human and Computer Vision, Neuroinformatics, Humanoids, Biologically motivated systems and artefacts Autonomous Systems, Linguistics, Sports Engineering, Computational Intelligence, Biosignal Processing, or Cognitive Materials as well as the methodologies behind them. Within the scope of the series are monographs, lecture notes, selected contributions from specialized conferences and workshops.

*Advisory Board*

Heinrich H. Bülthoff, MPI for Biological Cybernetics, Tübingen, Germany

Masayuki Inaba, The University of Tokyo, Japan

J.A. Scott Kelso, Florida Atlantic University, Boca Raton, FL, USA

Oussama Khatib, Stanford University, CA, USA

Yasuo Kuniyoshi, The University of Tokyo, Japan

Hiroshi G. Okuno, Kyoto University, Japan

Helge Ritter, University of Bielefeld, Germany

Giulio Sandini, University of Genova, Italy

Bruno Siciliano, University of Naples, Italy

Mark Steedman, University of Edinburgh, Scotland

Atsuo Takanishi, Waseda University, Tokyo, Japan

More information about this series at <http://www.springer.com/series/8354>

Aditya Joshi · Pushpak Bhattacharyya  
Mark J. Carman

# Investigations in Computational Sarcasm

 Springer

Aditya Joshi  
IITB-Monash Research Academy  
Indian Institute of Technology Bombay  
Mumbai, Maharashtra  
India

Mark J. Carman  
Faculty of Information Technology  
Monash University  
Melbourne, VIC  
Australia

Pushpak Bhattacharyya  
Department of Computer Science and  
Engineering  
Indian Institute of Technology Bombay  
Mumbai, Maharashtra  
India

ISSN 1867-4925

ISSN 1867-4933 (electronic)

Cognitive Systems Monographs

ISBN 978-981-10-8395-2

ISBN 978-981-10-8396-9 (eBook)

<https://doi.org/10.1007/978-981-10-8396-9>

Library of Congress Control Number: 2018932175

© Springer Nature Singapore Pte Ltd. 2018

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by Springer Nature

The registered company is Springer Nature Singapore Pte Ltd.

The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

# Preface

Sarcasm is defined as verbal irony that is intended to mock or ridicule. Existing sentiment analysis systems show a degraded performance in case of sarcastic text. Hence, computational sarcasm has received attention from the sentiment analysis community. Computational sarcasm refers to computational techniques that deal with sarcastic text. This monograph presents our investigations in computational sarcasm based on the linguistic notion of incongruity. For example, the sentence ‘*I love being ignored*’ is sarcastic because the positive word ‘*love*’ is incongruous with the negative phrase ‘*being ignored.*’ These investigations are divided into three parts: understanding the phenomenon of sarcasm, sarcasm detection, and sarcasm generation.

To first understand the phenomenon of sarcasm, we consider two components of sarcasm: implied negative sentiment and presence of a target. To understand how implied negative sentiment plays a role in sarcasm understanding, we present an annotation study which evaluates the quality of a sarcasm-labeled dataset created by non-native annotators. Following this, in order to show how the target of sarcasm is important to understand sarcasm, we first describe an annotation study which highlights the challenges in distinguishing between sarcasm and irony (since irony does not have a target while sarcasm does) and then present a computational approach that extracts the target of a sarcastic text.

We then present our approaches for sarcasm detection. To detect sarcasm, we capture incongruity in two ways: ‘*intra-textual incongruity*’ where we look at the incongruity within the text to be classified (*i.e.*, target text) and the ‘*context incongruity*’ where we incorporate information outside the target text. To detect incongruity within the target text, we present four approaches: (a) a classifier that captures sentiment incongruity using sentiment-based features (as in the case of ‘*I love being ignored*’), (b) a classifier that captures semantic incongruity (as in the case of ‘*A woman needs a man like a fish needs bicycle*’) using word embedding-based features, (c) a topic model that captures sentiment incongruity using sentiment distributions in the text (in order to discover sarcasm-prevalent topics such as work, college), and (d) an approach that captures incongruity in the language model using sentence completion. The approaches in (a) and

(c) incorporate sentiment incongruity relying on sentiment-bearing words, whereas approach in (b) and (d) tackles other forms of incongruity where sentiment-bearing words may not be present.

On the other hand, to detect sarcasm using contextual incongruity, we describe two approaches: (a) a rule-based approach that uses historical text by an author to detect sarcasm in the text generated by them and (b) a statistical approach that uses sequence labeling techniques for sarcasm detection in dialogue. The approach in (a) attempts to detect sarcasm that requires author-specific context, while that in (b) attempts to detect sarcasm that requires conversation-specific context. Finally, we present a technique for sarcasm generation. In this case, we use a template-based approach to synthesize incongruity and generate a sarcastic response to user input. The output of our sarcasm generation system obtains high scores on three quality parameters: coherence, grammaticality, and sarcastic nature. Also, the human evaluators are able to sufficiently identify the output of our system from that of a general purpose chatbot.

Our investigations demonstrate how evidences of incongruity (such as sentiment incongruity, semantic incongruity) can be modeled using different learning techniques (such as classifiers, topic models) for sarcasm detection and sarcasm generation. In addition, our findings establish the promise of novel problems, such as sarcasm target identification and sarcasm-versus-irony classification, and provide insights for future research in sarcasm detection.

# Contents

<b>1</b>	<b>Introduction</b>	1
1.1	Sentiment Analysis (SA)	2
1.1.1	Challenges	2
1.1.2	Research Problems	3
1.1.3	Applications	5
1.2	Sarcasm and Computational Sarcasm	5
1.3	Motivation	6
1.3.1	Turing Test-Completeness	7
1.3.2	Impact on Sentiment Classification	7
1.4	Prevalence of Sarcasm	8
1.4.1	In Popular Culture	9
1.4.2	On the Web	10
1.5	Sarcasm Studies in Linguistics	11
1.6	Incongruity for Sarcasm	14
1.7	Contribution	15
1.8	Literature Survey	16
1.8.1	Problem Definition	19
1.8.2	Datasets	19
1.8.3	Other Datasets (Dialogues, Syntactic Patterns, etc.)	21
1.8.4	Approaches	22
1.8.5	Rule-Based Approaches	22
1.8.6	Statistical Approaches	22
1.8.7	Deep Learning-Based Approaches	24
1.8.8	Shared Tasks and Benchmark Datasets	24
1.8.9	Reported Performance	24
1.8.10	Trends in Sarcasm Detection	26
1.8.11	Pattern Discovery	27
1.8.12	Role of Context in Sarcasm Detection	27
1.8.13	Issues in Sarcasm Detection	28

1.8.14	Issues with Annotation . . . . .	29
1.8.15	Issues with Sentiment as a Feature . . . . .	29
1.8.16	Dealing with Dataset Skews . . . . .	30
1.8.17	Sentiment Analysis at IIT Bombay . . . . .	30
1.8.18	Summary . . . . .	31
1.8.19	Monograph Organization . . . . .	31
<b>2</b>	<b>Understanding the Phenomenon of Sarcasm . . . . .</b>	<b>33</b>
2.1	Impact on Cross-Cultural Annotation . . . . .	33
2.1.1	Motivation . . . . .	34
2.1.2	Experiment Description . . . . .	35
2.1.3	Analysis . . . . .	36
2.2	Sarcasm-versus-irony Classification . . . . .	40
2.2.1	Motivation . . . . .	41
2.2.2	Experiment Description . . . . .	41
2.2.3	Analysis . . . . .	42
2.3	An Approach for Identification of the Sarcasm Target . . . . .	45
2.3.1	Motivation . . . . .	46
2.3.2	Architecture . . . . .	47
2.3.3	Experiment Description . . . . .	51
2.3.4	Results and Analysis . . . . .	53
2.4	Summary . . . . .	56
<b>3</b>	<b>Sarcasm Detection Using Incongruity Within Target Text . . . . .</b>	<b>59</b>
3.1	Sentiment Incongruity as Features . . . . .	59
3.1.1	Motivation . . . . .	60
3.1.2	Sentiment Incongruity-Based Features . . . . .	61
3.1.3	Experiment Setup . . . . .	63
3.1.4	Results . . . . .	64
3.1.5	Error Analysis . . . . .	65
3.2	Semantic Incongruity as Features . . . . .	66
3.2.1	Motivation . . . . .	67
3.2.2	Word Embedding-Based Features . . . . .	67
3.2.3	Experiment Setup . . . . .	69
3.2.4	Results . . . . .	70
3.2.5	Error Analysis . . . . .	73
3.3	Sentiment Incongruity Using Topic Model . . . . .	74
3.3.1	Motivation . . . . .	74
3.3.2	Model . . . . .	75
3.3.3	Experiment Setup . . . . .	78
3.3.4	Results . . . . .	78
3.3.5	Application to Sarcasm Detection . . . . .	83



3.4	Language Model Incongruity Using Sentence Completion . . . . .	84
3.4.1	Motivation . . . . .	85
3.4.2	Approach . . . . .	85
3.4.3	Experiment Setup . . . . .	87
3.4.4	Results . . . . .	88
3.4.5	Discussion . . . . .	89
3.4.6	Error Analysis . . . . .	90
3.5	Summary . . . . .	90
<b>4</b>	<b>Sarcasm Detection Using Contextual Incongruity . . . . .</b>	<b>93</b>
4.1	Contextual Incongruity in a Monologue . . . . .	93
4.1.1	Motivation . . . . .	94
4.1.2	Architecture . . . . .	95
4.1.3	Experiment Setup . . . . .	97
4.1.4	Results . . . . .	98
4.1.5	Error Analysis . . . . .	99
4.2	Contextual Incongruity in Dialogue . . . . .	100
4.2.1	Motivation . . . . .	101
4.2.2	Architecture . . . . .	101
4.2.3	Conversational Sarcasm Dataset . . . . .	103
4.2.4	Paradigm 1: Traditional Models . . . . .	106
4.2.5	Paradigm 2: Deep Learning-Based Models . . . . .	108
4.2.6	Results . . . . .	111
4.2.7	Error Analysis . . . . .	115
4.3	Summary . . . . .	117
<b>5</b>	<b>Sarcasm Generation . . . . .</b>	<b>119</b>
5.1	Motivation . . . . .	119
5.2	Architecture . . . . .	120
5.2.1	Input Analyzer . . . . .	120
5.2.2	Generator Selector . . . . .	122
5.2.3	Sarcasm Generator . . . . .	122
5.3	Evaluation . . . . .	125
5.3.1	Experiment Details . . . . .	125
5.3.2	Results . . . . .	126
5.4	Summary . . . . .	127
<b>6</b>	<b>Conclusion and Future Work . . . . .</b>	<b>129</b>
6.1	Summary . . . . .	129
6.2	Conclusion . . . . .	132
6.3	Future Work . . . . .	134
	<b>References . . . . .</b>	<b>137</b>

## About the Authors

**Aditya Joshi** successfully defended his Ph.D. thesis at IITB-Monash Research Academy, Mumbai, a joint Ph.D. program run by the Indian Institute of Technology Bombay (IIT Bombay) and Monash University, Australia, since January 2013. His primary research focus is computational sarcasm, and 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 natural language processing (NLP) such as sentiment analysis for Indian languages, drunk-texting prediction, news headline translation, and political issue extraction. The monograph is an outcome of his Ph.D. research.

**Dr. Pushpak Bhattacharyya** is the current president of the Association for Computational Linguistics (ACL) (2016–2017). He is the Director of the Indian Institute of Technology Patna (IIT Patna) and Vijay and Sita Vashee Chair Professor in the Department of Computer Science and Engineering at Indian Institute of Technology Bombay (IIT Bombay). He was educated at the Indian Institute of Technology Kharagpur (IIT Kharagpur) (B.Tech), Indian Institute of Technology Kanpur (IIT Kanpur) (M.Tech.), and IIT Bombay (Ph.D.).

He has been a Visiting Scholar and Faculty Member at the Massachusetts Institute of Technology (MIT), Stanford, UT-Houston, and University Joseph Fourier (France). His research areas include natural language processing, machine learning, and artificial intelligence (AI). Loved by his students for his inspiring teaching and mentorship, he has guided more than 250 students (Ph.D., masters, and bachelors). He has published over 250 research papers, is the author of the textbook ‘Machine Translation,’ and has led government and industry projects of international and national importance. His significant contributions in the field include multilingual lexical knowledge bases and projection. He is a fellow of the National Academy of Engineering and recipient of the IIT Bombay's Patwardhan Award and the Indian Institute of Technology Roorkee's (IIT Roorkee) VNMM award, both for technology development. He has also received IBM, Microsoft, Yahoo, and United Nations faculty grants.

**Dr. Mark J. Carman** is a Senior Lecturer at the Faculty of Information Technology, Monash University, Australia. He obtained a Ph.D. from the University of Trento, Italy, in 2004. His research and interests span from theoretical studies (e.g., investigating statistical properties of information retrieval measures) to practical applications (e.g., technology for assisting police during digital forensic investigations). He has authored a large number of publications in prestigious venues, including full papers at SIGIR, KDD, IJCAI, CIKM, WSDM, CoNLL, and ECIR and articles in TOIS, IR, JMLR, ML, PR, JAIR, and IP&M.