

C02029: Doctor of Philosophy
Subject Code: 32903
September 2022

CRICOS Code: 009469A

*Computational Understanding of Figurative
Language on Social Media*

Rhys Biddle

School of Computer Science
Faculty of Eng. & IT
University of Technology Sydney
NSW - 2007, Australia

Computational Understanding of Figurative Language on Social Media

*A thesis submitted in fulfilment of the requirements
for the degree of*

Doctor of Philosophy
in
Analytics

by

Rhys Biddle

to

Advanced Analytics Institute (AAi)
School of Computer Science
University of Technology Sydney
NSW - 2007, Australia

September 2022

CERTIFICATE OF ORIGINAL AUTHORSHIP

I, *Rhys Biddle* declare that this thesis, submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the School of Computer Science, Faculty of Engineering and Information Technology at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

Production Note:

SIGNATURE: Signature removed prior to publication.

[Rhys Biddle]

DATE: 7th September, 2022

PLACE: Melbourne, Australia

ABSTRACT

Figurative language in online user-generated text poses challenges to Natural Language Processing (NLP) systems designed to automate the understanding of natural language. This thesis introduces empirical studies that quantify the presence and describes the nature of figurative language in Social Media posts (i.e. Twitter). It also quantifies the impact of figurative language on particular NLP applications and introduces new resources (i.e. datasets and methodologies) for the computational processing of figurative language.

This thesis contains a focused case-study on general figurative language in the context of Public Health Surveillance (PHS) applications that monitor Twitter for health events. Findings indicate that some symptom and disease topics are mentioned in a figurative context more than in a health context, which results in a biased signal. To address this bias, a new annotated dataset and text classifier is proposed that reduces bias by targeting figurative expressions of health-related concepts on Twitter.

There is limited research on the expression of hyperbole on Twitter compared to other types of figurative language (e.g., metaphor). To address this gap, a dataset of tweets annotated for the presence of hyperbole is collected and explored. Findings show that hyperbole is relatively common on Twitter and the expression of hyperbole varies from simple and repetitive to complex and novel. A common theme of hyperbole expression on Twitter is the strong affective-laden intentions of the authors, heightening the importance of hyperbole understanding for affective computing applications. Several text classifiers are proposed that leverage pre-trained language models, affective signals, and privileged information for the detection of hyperbole. Experiments show improvements in the detection of hyperbole and importantly highlight annotation biases inherent in the current annotation scheme for hyperbole detection, which is likely to be a roadblock to further improvements.

This thesis quantifies the occurrence of figurative language on Twitter and demonstrates a considerable and consistent presence. Additionally, figurative language is often mishandled by various NLP resources and is scantily addressed by existing datasets and methodologies. Experiment results show that through direct targeting and careful handling of figurative language, improvements to the detection of figurative language are achievable. However, it is concluded that the complexity and novelty of figurative language requires further algorithmic and data inventions for continued progress.

ACKNOWLEDGMENTS

This thesis could not have been completed without the wisdom and guidance of many individuals from whom I sought advice throughout the journey. I acknowledge those who had a strong and direct influence on my thinking process and this thesis; Prof. Guandong Xu, Dr. Shaowu Liu, Dr. Aditya Joshi, Dr. Maciek Rybinksi and Dr. Cecile Paris. I also acknowledge my colleagues that helped me better formulate and articulate my ideas in our many conversations over this journey. To the anonymous reviewers that also provided insightful feedback to my submissions throughout my candidature I also extend my acknowledgements to you.

LIST OF PUBLICATIONS

RELATED TO THE THESIS :

- **Biddle, R.**, Joshi, A., Liu, S., Paris, C. and Xu, G., 2020, April. *Leveraging sentiment distributions to distinguish figurative from literal health reports on Twitter*. In Proceedings of The Web Conference 2020 (pp. 1217-1227).
- **Biddle, R.**, Rybinski, M., Li, Q., Paris, C. and Xu, G., 2021, December. *Harnessing Privileged Information for Hyperbole Detection*. In Proceedings of Australasian Language Technology Association (ALTA).

OTHERS :

- Islam, M.R., Liu, S., **Biddle, R.**, Razzak, I., Wang, X., Tilocca, P. and Xu, G., 2021. *Discovering dynamic adverse behavior of policyholders in the life insurance industry*. Technological Forecasting and Social Change, 163, p.120486.

NOMENCLATURE

- API Application Programming Interface
- BERT Bidirectional Encoder Representations from Transformers
- BiLSTM Bi-Directional Long Short Term Memory Network
- BNC British National Corpus
- CNN Convolutional Neural Network
- ECF Extreme Case Formulation
- ELMo Embeddings from Language Model
- GloVe Global Vectors for Word Representation
- GRNN Gated Recurrent Neural Network
- HMC Health Mention Classification
- ICD International Statistical Classification of Diseases and Related Health Problems
- ICF International Classification of Functioning, Disability and Health
- IR Information Retrieval
- KNN K-Nearest Neighbour
- LDA Latent Dirichlet Allocation
- LIWC Linguistic Inquirer Word Count
- LSTM Long Short Term Memory Network
- MELC Metaphor in end-of-life Care

MIP Metaphor Identification Procedure
NLG Natural Language Generation
NLP Natural Language Processing
NLU Natural Language Understanding
PHS Public Health Surveillance
RNN Recurrent Neural Network
RTT Round Trip Translation
SVM Support Vector Machine
TER Translation Edit Ratio
ULMFit Universal Language Model Fine Tuning for Text Classification
VAD Valence Arousal Dominance
w2v Word2Vec
WHO World Health Organization

TABLE OF CONTENTS

List of Publications	vii
Nomenclature	ix
List of Figures	xv
List of Tables	xvii
1 Introduction	1
1.1 Thesis Statement	1
1.2 Background	1
1.2.1 Figurative Language	2
1.2.2 Online User-Generated Content and Social Media	3
1.2.3 Natural Language Processing	4
1.3 Research Questions	6
1.4 Research Objectives	6
1.5 Thesis Contributions	7
1.5.1 Figurative Language and Public Health	7
1.5.2 Hyperbole	7
1.5.3 Towards Hyperbole Detection	8
1.5.4 Towards Hyperbole Interpretation	9
1.6 Thesis Structure	9
I Figurative Language and Public Health	11
2 Figurative Language and Health Mention Classification	13
2.1 Introduction	13
2.2 Figurative Language and Public Health	14

TABLE OF CONTENTS

2.2.1	Public Health Surveillance	16
2.2.2	Epidemic Intelligence	17
2.3	HMC2019	20
2.4	Health Mention Classification and Figurative Language Bias	23
2.5	Word Representations for Health Mention Classification	26
2.6	Health Mention Classification	28
2.6.1	Baselines	29
2.6.2	BiLSTM+Senti	30
2.6.3	Experimental Setup	34
2.7	Results & Discussion	35
2.8	Error Analysis	37
2.9	Conclusion	38
 II Hyperbole		41
 3 Hyperbole		43
3.1	Introduction	43
3.2	Hyperbole as Important Figure of Speech	44
3.2.1	Hyperbole	45
3.2.2	Hyperbole Types	46
3.3	HYPO	49
3.4	HyperTwit	50
3.4.1	Data Collection	50
3.4.2	Annotation and Inter-Annotator Agreement Study	52
3.5	Hyperbole on Twitter	57
3.5.1	Hyperbole Prevalence	57
3.5.2	Word Hyperbolicity	58
3.5.3	Common Intentions	59
3.5.4	Diversity of Hyperbole	64
3.6	HyperProbe	65
3.6.1	Extreme Case Formulation Tests	65
3.6.2	Qualitative Hyperbole Tests	69
3.6.3	Quantitative Hyperbole Tests	70
3.7	Conclusion	74

4	Towards Computational Hyperbole Detection	77
4.1	Introduction	77
4.2	Motivation and Methods for Hyperbole Detection	78
4.2.1	Motivation for Hyperbole Detection	78
4.2.2	Methods for Figurative Language Detection	79
4.3	Methodology	83
4.3.1	Baselines	83
4.3.2	Affective Signals for Hyperbole Detection	85
4.3.3	Privileged Information for Hyperbole Detection	90
4.4	Experiments and Results	97
4.4.1	In-Domain Hyperbole Detection	98
4.4.2	Cross-Domain Hyperbole Detection	103
4.4.3	HyperProbe Experiments	105
4.5	Conclusion	112
5	Towards Computational Hyperbole Interpretation	117
5.1	Introduction	117
5.2	Natural Language Generation and Figurative Language	118
5.3	Methodology	121
5.3.1	Hyperbole Interpretation	121
5.3.2	Baselines and Models	122
5.4	Experiments	124
5.5	Results	127
5.6	Error Analysis	129
5.7	Conclusion	130
III	Conclusion	131
6	Discussion and Conclusion	133
6.1	Introduction	133
6.2	Thesis Statement	133
6.3	Research Questions	133
6.3.1	Research Question i)	134
6.3.2	Research Question ii)	134
6.3.3	Research Question iii)	135

TABLE OF CONTENTS

6.4	Research Objectives	136
6.4.1	Research Objective i)	136
6.4.2	Research Objective ii)	136
6.4.3	Research Objective iii)	137
6.5	Future Directions	138
6.5.1	Hyperbole	138
6.5.2	Metaphor and Sarcasm	140
	Bibliography	143

LIST OF FIGURES

FIGURE	Page
1.1 Figurative Language Utterances	3
2.1 Typical Epidemic Intelligence Pipeline	17
2.2 Correlation Matrix Heatmap	24
2.3 BiLSTM + Senti	33
3.1 Example Data	52
3.2 Annotator Interpretation Similarity Distributions	55
3.3 Daily Proportion of Hyperbole	58
3.4 Hyperbole Cluster Summary Heatmap	63
4.1 LR+QQ Model Diagram	83
4.2 BERT+QQ Model Diagram	84
4.3 BERT+3dEmo	87
4.4 BERT+3dEmoMT	87
4.5 BERT+3dEmoAS	89
4.6 Diagram of Hyperbole and Literal Paraphrase Examples	91
4.7 BERT+PI	93
4.8 Triplet Sampling Example	93
4.9 Model Explanation Comparisons - HYPO	100
4.10 Model Explanation Comparisons - HYPO	100
4.11 LIME Explanations - HyperProbe (ECFs)	106
4.12 LIME Explanations - HyperProbe (Qualitative Hyperbole)	107
4.13 LIME Explanations - HyperProbe (Qualitative Hyperbole)	107
4.14 LIME Explanations - HyperProbe (Qualitative Hyperbole)	108
4.15 LIME Explanations - HyperProbe (Quantative Dimensions)	110
4.16 Model-Experiment Rankings (F1)	113

4.17 Experiment-Model Rankings (F1)	114
5.1 Example Data	118
5.2 Diagram of Tagging Module from Modular	123
5.3 Diagram of Modular	123

LIST OF TABLES

TABLE	Page
1.1 The landscape of NLP	5
2.1 HMC2019 Statistics	20
2.2 HMC2019 Examples	21
2.3 Baseline HMC results	23
2.4 Word-Representations and Health Mention Classification	28
2.5 Results of HMC experiments (BERT)	35
2.6 Results of HMC experiments (ELMo)	35
2.7 Results of HMC experiments (w2v)	36
2.8 Results by Keyword	36
3.1 Hyperbole Types	47
3.2 Overview of Dataset Contributions in this Thesis	48
3.3 HYPO examples	49
3.4 Hyperbole term list	53
3.5 Hyperbole Types - HyperTwit	54
3.6 Top 15 words by Removal Probability - HyperTwit _R	60
3.7 Top 15 words by Removal Probability - HyperTwit _K	60
3.8 Top 15 words Added to Literal Interpretations - HyperTwit _R	61
3.9 Top 15 words Added to Literal Interpretations - HyperTwit _K	62
3.10 Hyperbolic Expressions of ‘blind’	62
3.11 Hyperbolic Expressions of ‘toxic’	63
3.12 HyperProbe Test Statistics.	66
3.13 ECF Adjectives Test Sentence Templates	67
3.14 ECF Adverbs Test Sentence Templates	67
3.15 ECF Determiners Test Sentence Templates	68
3.16 ECF Indefinite Pronouns Test	69

3.17	Qualitative Adjectives Test Sentence Templates	70
3.18	Quantitative Dimensions Test Sentence Templates	71
3.19	Quantitative Time Periods Test Sentence Templates	72
3.20	Quantitative Time Periods (Numeric) Test Sentence Templates	73
3.21	DoQ-Intrinsic Sentence Templates	74
3.22	DoQ-Intrinsic Test Sentence Example	74
4.1	Semi-Random Triplet Sample Examples	95
4.2	Similarity Triplet Sample Examples	97
4.3	Hyperparameter search	97
4.4	Results from In-Domain Experiments - HYPO	99
4.5	Results from In-Domain Experiments - HyperTwit_K	99
4.6	Keyword Results by Hyperbole Type on HyperTwit_k	101
4.7	Cross-Domain Results - Idiomatic Hyperbole Evaluation	104
4.8	Cross-Domain Results - Twitter Hyperbole Evaluation	104
4.9	HyperProbe Results. Extreme Case Formulations	106
4.10	HyperProbe Results (Qualitative Hyperbole)	106
4.11	Hyperprobe Results. Quantitative Dimensions	109
4.12	HyperProbe Results. Time Periods	111
4.13	HyperProbe Results. Intrinsic Quantities	111
5.1	Paraphrase Experiment Results	126
5.2	Paraphrase Experiment Results - Manual Evaluation	126
5.3	Paraphrase Errors 1	128
5.4	Paraphrase Errors 2	128
5.5	Hyperbolicity, Hyperbole Type and Model	129