

CONTENTS

1	INTRODUCTION	1
1.1	Motivation	4
1.1.1	Sentiment Analysis of Twitter: Gaps and Challenges	5
1.1.2	From Affect Words to Words' Semantics	8
1.2	Research Questions, Hypotheses and Contributions	10
1.3	Thesis Methodology and Outline	16
1.4	Publications	20
i	BACKGROUND	23
2	LITERATURE REVIEW	25
2.1	Background	26
2.1.1	Fundamentals	26
2.1.2	A Note on Terminology	33
2.2	Sentiment Analysis of Twitter	35
2.2.1	Traditional Sentiment Analysis Approaches	36
2.2.1.1	The Machine Learning Approach	36
2.2.1.2	The Lexicon-based Approach	51
2.2.1.3	The Hybrid Approach	59
2.2.1.4	Discussion	62
2.3	Semantic Sentiment Analysis	65
2.3.1	Contextual Semantics	65
2.3.2	Conceptual Semantics	70
		74
2.4	Summary and Discussion	78

2.4.1	Discussion	82
ii	SEMANTIC SENTIMENT ANALYSIS OF TWITTER	87
3	CONTEXTUAL SEMANTICS FOR SENTIMENT ANALYSIS OF TWITTER	89
3.1	Introduction	89
3.2	The SentiCircle Representation of Words' Semantics	92
3.2.1	Overview	92
3.2.2	SentiCircle Construction Pipeline	94
3.2.2.1	Term Indexing	95
3.2.2.2	Context Vector Generation	96
3.2.2.3	SentiCircle Generation	97
3.2.2.4	Senti-Median: The Overall Contextual Sentiment Value	101
3.3	SentiCircles for Sentiment Analysis	103
3.3.1	Entity-level Sentiment Detection	103
3.3.2	Tweet-level Sentiment Detection	103
3.3.2.1	The Median Method	104
3.3.2.2	The Pivot Method	104
3.3.2.3	The Pivot-Hybrid Method	106
3.3.3	Evaluation Setup	106
3.3.3.1	Datasets	106
3.3.3.2	Sentiment Lexicons	111
3.3.3.3	Baselines	111
3.3.3.4	Thresholds and Parameters Tuning	112
3.3.4	Evaluation Results	114
3.3.4.1	Entity-Level Sentiment Detection	114
3.3.4.2	Tweet-Level Sentiment Detection	116

3.3.4.3	Impact on Words' Sentiment	119
3.4	SentiCircles for Adapting Sentiment Lexicons	120
3.4.1	Evaluating SentiStrength on the Adapted Thelwall-Lexicon	123
3.5	Runtime Analysis	128
3.6	Discussion	129
3.7	Summary	132
4	CONCEPTUAL SEMANTICS FOR SENTIMENT ANALYSIS OF TWITTER	135
4.1	Introduction	135
4.2	Conceptual Semantics for Supervised Sentiment Analysis	138
4.2.1	Extracting Conceptual Semantics	140
4.2.2	Conceptual Semantics Incorporation	141
4.2.3	Evaluation Setup	144
4.2.3.1	Datasets	144
4.2.3.2	Semantic Concepts Extraction	145
4.2.3.3	Baselines	147
4.2.4	Evaluation Results	150
4.2.4.1	Results on Incorporating Semantic Features	151
4.2.4.2	Comparison of Results	152
4.3	Conceptual Semantics for Lexicon-based Sentiment Analysis	156
4.3.1	Enriching SentiCircles with Conceptual Semantics	157
4.3.2	Evaluation Results	158
4.4	Discussion	160
4.5	Summary	162

5	SEMANTIC PATTERNS FOR SENTIMENT ANALYSIS OF TWITTER	165
5.1	Introduction	165
5.2	Related Work	167
5.3	Semantic Sentiment Patterns of Words	169
5.3.1	Syntactical Preprocessing	169
5.3.2	Capturing Contextual Semantics and Sentiment of Words	170
5.3.3	Extracting Patterns from SentiCircles	171
5.4	Evaluation Setup	173
5.4.1	Tweet-Level Evaluation Setup	173
5.4.2	Entity-Level Evaluation Setup	174
5.4.3	Evaluation Baselines	177
5.4.4	Number of SS-Patterns in Data	179
5.5	Evaluation Results	180
5.5.1	Results of Tweet-Level Sentiment Classification	181
5.5.2	Results of Entity-Level Sentiment Classification	183
5.6	Within-Pattern Sentiment Consistency	187
5.6.1	Sentiment Consistency vs. Sentiment Dispersion	188
5.7	Discussion	190
5.8	Summary	192
iii	ANALYSIS STUDY	195
6	STOPWORD REMOVAL FOR TWITTER SENTIMENT ANALYSIS	197
6.1	Introduction	198
6.2	Stopword Analysis Set-Up	200
6.2.1	Datasets	201
6.2.2	Stopword removal methods	201

6.2.2.1	The Classic Method	202
6.2.2.2	Methods based on Zipf's Law (<i>Z-Methods</i>)	202
6.2.2.3	Term Based Random Sampling (<i>TBR</i> <i>S</i>)	203
6.2.2.4	The Mutual Information Method (<i>MI</i>)	204
6.2.3	Twitter Sentiment Classifiers	205
6.3	Evaluation Results	206
6.3.1	Classification Performance	206
6.3.2	Feature Space	209
6.3.3	Data Sparsity	210
6.3.4	The Ideal Stoplist	213
6.4	Discussion	215
6.5	Summary	216
iv	CONCLUSION	219
7	DISCUSSION AND FUTURE WORK	221
7.1	Discussion	222
7.1.1	Extracting Words' Semantics	222
7.1.1.1	Extracting Contextual Semantics	223
7.1.1.2	Extracting Conceptual Semantics	224
7.1.1.3	Extracting Stopwords	225
7.1.2	Incorporating Words' Semantics in Sentiment Analysis	226
7.1.2.1	Incorporating Semantics into Lexicon-based Approaches	226
7.1.2.2	Incorporating Semantics into Machine Learning Approaches	228
7.1.2.3	Words' Semantics for Adapting Sentiment Lexicons	229

7.1.3	Assessment and Results	230
7.2	Future Work	232
8	CONCLUSION	239
8.1	Contextual Semantics for Sentiment Analysis of Twitter	241
8.2	Conceptual Semantics for Sentiment Analysis of Twitter	242
8.3	Semantic Patterns for Sentiment Analysis of Twitter	243
8.4	Analysis on Stopword Removal Methods for Sentiment Analysis of Twitter	245
V	APPENDIX	247
A	EVALUATION DATASETS FOR TWITTER SENTIMENT ANALYSIS	249
B	ANNOTATION BOOKLET FOR THE STS-GOLD DATASET	257
	BIBLIOGRAPHY	259