

Contents

1	Introduction	1
1.1	Objectives	3
1.2	Why Twitter?	4
1.3	Applications	4
1.4	Thesis statement	4
1.5	Challenges	5
1.6	Contributions	6
2	Background	8
2.1	Twitter and Social Media	8
2.1.1	Twitter Platform	8
2.1.2	Twitter API	9
2.2	Supervised learning: Classification	9
2.3	Unsupervised learning: Clustering	13
2.4	Neural Network-based Language Models	17
3	Related Work	20
3.1	Common Tasks involving Events	20
3.1.1	Event Detection	20
3.1.2	Automatic Summarization of Events	21
3.1.3	Systems for Event Exploration	22
3.1.4	Improving News Delivery from Social Media	22
3.2	Event Models	23
3.2.1	Modeling the Impact of Events	23
3.2.2	Location-aware Event Models	25
3.2.3	Models of Aggregated Event Content	26
3.3	Quantitative Analysis of News Events	27
4	Data Collection Methodology	29
4.1	Building the Dataset	29
4.1.1	Collecting Related Posts for News on Twitter	29
4.1.2	Identifying Events	32
4.2	Data Cleaning and Validation	35
4.2.1	Detecting Articulation Words	35
4.2.2	Validation of Event Modeling	37
4.2.3	Event Duration	39

5	User Reaction: Predicting and Characterizing High-Activity News Events	41
5.1	Introduction	41
5.2	The VQ-Event Model	42
5.3	Experimental Analysis	44
5.4	Results and Discussion	46
5.4.1	Information Forwarding Characteristics	49
5.4.2	Conversational Characteristics	51
5.4.3	Topical Focus Characteristics	51
5.4.4	Predicting the Activity of Events	52
5.5	Chapter Summary and Conclusions	54
6	Spatio-Temporal Context: Discovering Geopolitical Interactions on Twitter	56
6.1	Introduction	56
6.2	Event Representation	58
6.2.1	Event Representation Definition	59
6.2.2	Representing Relations Among Locations	60
6.2.3	System Architecture	62
6.3	Exploratory Analysis	64
6.4	Limitations	75
6.5	Chapter Summary and Conclusions	75
7	Aggregating Content: A Lightweight Model of News Events	77
7.1	Introduction	77
7.2	Event Representation	79
7.3	Case Studies	80
7.3.1	Datasets	80
7.3.2	Experimental setting	82
7.3.3	Validation of Sub-Topic Detection Task	83
7.4	Chapter Summary and Conclusions	85
8	Conclusions	88
8.1	Summary of Contributions	88
8.2	Limitations and Future Directions	89
	Bibliography	92
	Appendix A List of News Sources for Data Collection	104