

Social Media Text Mining for Decision Support in Natural Disaster Management in Sri Lanka

Submitted by:

K.H.J.Imalka

158761R

Supervised by: Mr. Saminda Premaratne

Faculty of Information Technology

University of Moratuwa

May 2018

**Social Media Text Mining for Decision Support in Natural
Disaster Management in Sri Lanka**

Submitted by:

K.H.J.Imalka

158761R

Supervised by: Mr. Saminda Premaratne

Dissertation submitted to the Faculty of Information Technology, University of
Moratuwa, Sri Lanka for the partial fulfillment of the requirements of the Master Degree
of Science in Information Technology.

May 2018

Declaration

I declare that this is my own work and has not been submitted in any form for another degree or diploma at any university or other institution of tertiary education. Information derived from the published or unpublished work of others has been acknowledged in the text and a list of reference is given.

Name of the student : K. H.J. Imalka

Student Number : 158761R

Signature of the student : Date:

Supervised by:

Name of the supervisor : Mr. Saminda Premarathne

Signature of the supervisor : Date:

Dedication

This Dissertation is dedicated to my loving parents and husband for being part of me and encouraging me always by being my side.

Acknowledgement

First I express my heartfelt gratitude to my supervisor Mr. Saminda Premarathna for his most valued guidance, commitment and kind support to make this research a success.

Also I sincerely thank Prof. Asoka S Karunananda who taught us all the research related document preparation which was a great support to manage all work with busy schedules.

It's my pleasure to thank Mr. Chaman Wijesiriwardena and all other Senior Lecturers, Lecturers, Instructors and staff members who helped us in many ways to successfully complete this research.

Then I would like to thank all the batch mates of MSc. In Information Technology batch 9 for their companionship and various kinds of support given throughout the program.

In addition, I would like to thank my work mates and friends for encouraging me with their support and best wishes.

Last but not least, my sincere thank goes to my beloved parents, husband and his parents for helping me to conduct this work without much stress and encouraging me to complete this research.

Abstract

With the popularity of internet and smart devices, social media is very popular today among individuals in almost all the ages which help them to create and share their personal feelings, experiences, ideas as well as information with others connected to them over a computer mediated technologies. Individuals use these social media applications such as Facebook and twitter which are popular most to share their experiences, opinions, day today activities as well as achievements. Due to this nature when there are emergencies and natural disasters these social media applications tend to be flooded with content generated from public who affected, who are looking for their family members and friends, who are looking for information as well as with the people engage in humanitarian activities.

Therefore social media has become the first to generate related information when there is a catastrophic event before any of news sites or government bodies engage in disaster management. These social media content is quick accurate and subjective during disaster situations therefore we can use this information as an asset to reduce risk and build awareness among public about the disaster as well as to provide decision making support to relief efforts.

This research focuses on building decision making support using social media content generated during disaster situations in Sri Lankan context. Mainly the content will be tweets posted by public during a natural disaster and consisting with text written in English. Therefore situational awareness building will be done using text mining which natural language processing in this study since the content is unstructured.

Content will be analyzed using techniques to scrape, clean, classify and generate real information about the disaster and to visualize them to support decision making for authorities engage in disaster management as well as volunteers engage in relief efforts.

Contents

DECLARATION.....	III
DEDICATION	II
ACKNOWLEDGEMENT	III
ABSTRACT	IV
TABLE OF FIGURES.....	VIII
TABLE OF TABLES	VIII
CHAPTER 1.....	1
INTRODUCTION TO SOCIAL MEDIA TEXT MINING FOR DECISION SUPPORT IN NATURAL DISASTER MANAGEMENT	1
1.1 INTRODUCTION	1
1.2 BACKGROUND OF STUDY	3
1.2.1 Disasters and Disaster Communication	3
1.2.2 Social Media.....	3
1.3 MOTIVATION	5
1.4 PROBLEM IN BRIEF.....	7
1.5 PROPOSED SOLUTION	8
1.6 AIMS AND OBJECTIVES	8
CHAPTER 2.....	10
LITERATURE REVIEW ON SOCIAL MEDIA TEXT MINING & DECISION MAKING	10
2.1 INTRODUCTION	10
2.2 USE OF SOCIAL MEDIA DURING DISASTER SITUATIONS	11
2.3 ASSESSMENT OF DISASTERS, PROCESSING SOCIAL MEDIA MESSAGES AND ASSIST DECISION MAKING IN MASS EMERGENCIES	12
2.4 OPPORTUNITIES AND BARRIERS OF USING SOCIAL MEDIA FOR DISASTER PREPAREDNESS	123
2.5 PRACTICAL EXTRACTION OF RELEVANT INFORMATION.....	104
2.6 SUMMARY	105

CHAPTER 3.....	16
TECHNOLOGY ADAPTED IN SOCIAL MEDIA TEXT MINING & DECISION MAKING	16
3.1 INTRODUCTION	16
3.2 TECHNOLOGIES TO COLLECT DATA IN SOCIAL MEDIA.....	16
3.2.1 <i>Twitter API</i>	18
3.2.2 <i>Facebook Graph API</i>	18
3.3 TEXT MINING	18
3.4 SENTIMENT ANALYSIS.....	19
3.5 CONTENT ANALYSIS FOR DECISION MAKING.....	20
3.6 SCIENTIFIC PROGRAMMING TOOLS	21
3.7 BUSINESS TOOLKITS.....	21
3.8 TEXT MINING CAPABILITIES OF RAPIDMINER	22
3.9 SUMMARY	22
CHAPTER 4.....	23
APPROACH FOR SOCIAL MEDIA TEXT MINING IN DISASTER SITUATIONS...23	
4.1 INTRODUCTION	23
4.2 HYPOTHESIS	23
4.3 INPUT.....	23
4.4 OUTPUT.....	23
4.5 PROCESS.....	23
4.6 INTRODUCTION TO DESIGN.....	24
4.7 TOP LEVEL DESIGN OF THE SYSTEM	24
4.8 SUMMARY	226
CHAPTER 5.....	25
ANALYSIS AND DESIGN OF PROPOSED SOLUTION.....	26
5.1 INTRODUCTION	26
5.2 SYSTEM DESIGN	26
5.3 RAPIDMINER PROCESS MODEL	27
5.3.1 <i>Text Processing</i>	27
5.3.2 <i>RapidMiner Text Processing Extension Package</i>	27
5.4 CLASSIFIER - .NET TECHNOLOGY BASED TOOL.....	27
5.4.1 <i>C#</i>	28
5.4.2 <i>Classifier Design</i>	28
CHAPTER 6.....	29

IMPLEMENTATION	29
6.1 INTRODUCTION	29
6.2 CHALLENGES IN PROPOSED SYSTEM IMPLEMENTATION	29
6.3 DOWNLOADING TWEETS USING TWEETS RETRIEVAL TOOL (V1.2).....	30
6.4 LABELED DATA	32
6.5 ANNOTATION SCHEME	32
6.6 ASSOCIATION RULE MINING	34
6.6.1 Text processing step by step	35
6.7 CLUSTERING TWEETS USING ASSOCIATION RULES	38
6.8 CLASSIFIER	38
6.9 SUMMARY	38
CHAPTER 7.....	42
EVALUATION	42
7.1 INTRODUCTION	42
7.2 EVALUATION OF ASSOCIATION RULE MINING.....	42
7.3 EVALUATION OF DIFFERENT CLASSIFIERS.....	41
7.4 SUMMARY.....	41
CHAPTER 8.....	47
DISCUSSION	47
8.1 INTRODUCTION	47
8.2 LIMITATIONS	48
8.3 FUTURE WORK	48
8.4 SUMMARY	49
REFERENCES.....	50
Appendix A – Association Rules.....	52

Table of Figures

Figure 1.1: Global Digital Snapshot	5
Figure 1.2: Active Users of Key Global Social Platforms	6
Figure 2.1 : Characteristic features of Twitter activity across locations	12
Figure 3.1: Text mining process	19
Figure 3.2: Machine Learning Overview	21
Figure 4.1: Approach to analyze data	24
Figure 4.2: Top Level Design of the System	24
Figure 5.1: Design of Propose System	26
Figure 6.1: shows how the data downloaded through Twitter Streaming API using the Tweets Retrieval Tool v1.2	31
Figure 6.2: Sample of full texts of tweets downloaded	32
Figure 6.3: Annotated tweets by paid workers from the Crowdfower crowdsourcing platform for labeling	33
Figure 6.4: Annotated data after loading to RapidMiner with Category and Tweet-text attributes	34
Figure 6.5: Operators for Process and the Parameters for Process Documents from Files operator	34
Figure 6.6: Operators within the Process Documents from Files nested operator	35
Figure 6.7: Process Documents from Files sub-process output	37
Figure 6.8: Table View for the AssociationRules generated by the Process	38
Figure 6.9: Using write operator to write association rules into XML	39
Figure 6.10: XML file of Association Rules	39
Figure 6.11 Main interface of Classifier	40
Figure 6.12: Categorized tweets	41
Figure 7.1: Parameters of Create Association Rules Operator	42
Figure 7.2 : Data (Tabular) view Association Rules	43
Figure 7.3: Graph view of Association Rules	43
Figure 7.4: Association Rules of Caution and Advice Category	44
Figure 7.5 – Performance of K-NN algorithm	45
Figure 7.6 – Performance of Naïve Bayes algorithm	43

Table of Tables

Table 1: Social Media Types with Examples	4
Table 2: Applications, Modules and Libraries to collect data from Facebook and Twitter	17
Table 7.1 Accuracy of different classifiers	46