UNIVERSITI PUTRA MALAYSIA

SARCASM DETECTION MODEL BASED ON TWEETS' STRENGTH USING HASHTAGS AND NON-HASHTAGS SENTIMENT ANALYSIS

SAMANEH NADALI

FSKTM 2016 46

SARCASM DETECTION MODEL BASED ON TWEETS' STRENGTH USING HASHTAGS AND NON-HASHTAGS SENTIMENT ANALYSIS

By

SAMANEH NADALI

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfilment of the Requirements for the Degree of Doctor of Philosophy

July 2016

DEDICATIONS

You step into the path and never ask
The path itself will tell you how to get to the end
Farid ud-Din Attār
Persian poet
(c.1110-c.1221)

To may Beloved Parents
Tayebeh & Mohammad Hossein

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirement for the degree of Doctor of Philosophy

SARCASM DETECTION MODEL BASED ON TWEETS´ STRENGTH USING HASHTAGS AND NON-HASHTAGS SENTIMENT ANALYSIS

By

SAMANEH NADALI

July 2016

Chairman:Masrah Azrifah Azmi Murad, PhD
Faculty: Computer Science and Information Technology

Recently, microblogs platforms such as Twitter are becoming popular day by day. People used Twitter for building common ground, sharing information and sharing opinions on a variety of topics and discussing current issues. Thus, Twitter becomes source of opinions. Therefore understanding the sentiment of the opinion is needed.

Over the last decades, sentiment analysis (SA) in social media has been one of the most research areas in Natural Language Processing (NLP). The aim of sentiment analysis is to automatically identify the polarity of a document, where misinterpreting irony and sarcasm is a big challenge. There is a weak boundary in the meaning between irony, sarcasm and satire, therefore in this thesis only the term sarcasm is employed.

Sarcasm is a common phenomenon in social media, which is a nuance form of language for expressing the opposite of what is inferred. Sarcasm generally changes the polarity of an utterance from positive or negative into its opposite. Therefore, identifying sarcasm correctly can enhance the performance of sentiment classification. Sarcasm analysis is a difficult task not only for the machine, but also for a human, because of the intentional ambiguity. Although sarcasm detection has an important effect on sentiment, it is usually ignored in social media analysis because sarcasm analysis is too complicated.

Several techniques have been used in sarcasm detection such a semi-supervised, detection sarcasm based on intensifiers and exclamation, the impact of lexical

and pragmatic factors, contrast between positive and negative situation verb phrases and hashtags based sentiment analysis. In this thesis, two existing works; sarcasm as a contrast between positive sentiment and negative situation phrases and hashtags based sentiment analysis are extended. For the former task, the authors of the work have presented a novel bootstrapping algorithm that automatically learns a list of positive sentiment phrases and negative situation phrases from sarcastic tweets. The results showed a contrast between positive and negative and they can be used in recognizing sarcastic tweets. However, the work only identified one type of sarcasm tweets (i.e. positive verb phrases followed by negative situation phrases). In additional they did not work on identifying sarcasm when a negative situation phrases is followed by positive sentiment in the separate sentences. Moreover, the intensity of the negativity is not considered in their work. In addition, the work did not consider hashtags and sentiment analysis of hashtags. Hashtag is a topic or key words that are marked with a tweet. Since many of the hashtags contain polarity, detection of sarcasm at hashtags level will have a positive effect on polarity classification.

The later work which is extended in this thesis works based on the hashtags sentiment analysis. The authors identified sarcastic tweets based on the sarcasm indicators and contrast between the sentiment orientation of the tweets and hashtags. Although, the work was primary work at the level of the hashtags sentiment analysis, they did not use systematic approach for identifying sarcasm indicators. Moreover, they worked only based on the contrast between the sentiment orientation of the tweets and hashtags. Since sarcasm utterance contains hyperbole and exaggeration and some hashtags are used for emphasizing the text, identifying based on the contrast between the sentiment of the tweets and hashtags is not sufficient.

To address problems, a Sarcasm Detection Model (SDM) is proposed. In the proposed model, three classifiers; SentiStrength Sarcasm Classifier (SSC), Sarcasm Hashtags Classifier (SHC) and Hashtags-SentiStrength Sarcasm Classifier (HSSC) is used. SSC is worked at the level of the non-hashtags sentiment analysis, whereas SHC and HSSC at the level of the hashtags sentiment analysis. In the SSC, sarcasm is identified based on the strength level of tweets. Several lexical and pragmatic features such as emoticons, interjections, capital words and elongate words are applied in the proposed SentiStrength formula.

Sarcasm Hashtags Classifier (SHC) is used to identify sarcastic tweets based on the Sarcasm Hashtags Indicator (SHI) and Sentiment Hashtags Analysis (SHA). In the classifier (SHC), a bootstrapping algorithm is used to identify Sarcasm Hashtags Indicator (SHI). SHI contains a list of hashtags that help to identify sarcastic tweets easily. In the proposed model (SDM), if a tweet contains SHI, it will be labeled as sarcastic tweet; otherwise the Sentiment Hashtags Analysis (SHA) is applied. SHA is worked based on the contrast between sentiment orientation of the tweets and hashtags. In this part, the hashtags are retokenized through preprocessing and the orientation of the hashtags is identified. Next, the orientation of a tweet without hashtags is also identified. The tweet is considered as sarcasm hashtags if there is a contrast between the orientation of the tweet and hashtags.

The HSSC, works based on the strength level of tweets and hashtags. In this classifier, the effect of the sentiment of the hashtags for increasing the polarity of the tweets is considered.

The Sarcasm Detection Model (SDM) has been tested on two datasets which each dataset contains 3000 sarcastic and non- sarcastic tweets. All of the tweets were extracted randomly using the Twitter API. So far, no work has been done in sarcasm detection at the level of hashtags and non-hashtags based sentiment analysis. So, the novelty of the proposed model (SDM) is in identifying sarcastic tweets by analyzing strength of the tweets at the level of the hashtags and non-hashtags sentiment analysis. The results of the study (0.85% of precision) demonstrates that the SDM is more accurate and effective than the existing works which was done based on the contrast between positive and negative situation phrases and hashtags based sentiment analysis.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Doktor Falsafah

MODEL PENGESANAN SINDIRAN BERDASARKAN KEKUATAN TWEET MENGGUNAKAN ANALISIS SENTIMEN HASHTAGS DAN BUKAN HASHTAGS

Oleh

SAMANEH NADALI

Julai 2016

Pengerusi: Masrah Azrifah Azmi Murad, PhD
Fakulti: Sains Komputer dan Teknolologi Maklumat

Baru-baru ini, platform mikroblog seperti Twitter menjadi popular hari demi hari. Orang awam menggunakan Twitter untuk membina persefahaman, perkongsian maklumat dan berkongsi pendapat mengenai pelbagai topik dan membincangkan isu-isu semasa. Oleh itu, Twitter merupakan sumber pendapat. Maka, memahami sentimen bagi pendapat amat diperlukan.

Sejak beberapa dekad yang lalu, analisis sentimen (SA) dalam media sosial telah menjadi salah satu bidang penyelidikan paling utama dalam Pemprosesan Bahasa Asli (NLP). Tujuan analisis sentimen adalah untuk mengenal pasti kekutuban sesuatu dokumen, di mana mentafsirkan ironi dan sindiran secara otomatik adalah satu cabaran besar. Terdapat sempadan lemah dalam makna di antara ironi, sindiran dan satira, oleh itu dalam tesis ini hanya sindiran istilah ini digunakan.

Sindiran adalah satu fenomena biasa dalam media sosial, yang merupakan satu bentuk nuansa bahasa bagi menyatakan yang bertentangan dengan apa yang dibayangkan. Sindiran umumnya menukarkan kekutuban sesuatu ujaran daripada positif atau negatif ke ujaran bertentangan. Oleh itu, mengenal pasti sindiran dengan betul boleh meningkatkan prestasi klasifikasi sentimen. Analisis sindiran adalah tugas yang sukar bukan sahaja untuk mesin, tetapi juga untuk manusia, kerana kekaburan yang disengajakan. Walaupun pengesanan sindiran mempunyai kesan penting kepada sentimen, ia biasanya diabaikan dalam analisis media sosial kerana analisis sindiran terlalu rumit.

Beberapa teknik telah digunakan dalam pengesanan sindiran seperti separuh seliaan, sindiran pengesanan berdasarkan penguat dan seru, kesan faktor leksikal dan pragmatik, perbezaan antara frasa kata kerja positif dan negatif dan analisis sentimen berasaskan *hashtag*. Dalam tesis ini, dua kerja penyelidikan sedia ada iaitu sindiran sebagai kontra antara sentimen positif dan frasa keadaan negatif dan analisis sentimen berdasarkan *hashtag* diperluaskan. Bagi kerja sedia ada yang pertama, penyelidik telah membentangkan algoritma bootstrapping baharu yang secara otomatik mempelajari senarai frasa sentimen positif dan frasa keadaan negatif daripada *tweet* sindiran. Hasil kajian menunjukkan satu perbezaan antara positif dan negatif dan ianya boleh digunakan dalam mengiktiraf *tweet* sindiran. Walau bagaimanapun, kerja-kerja penyelidikan hanya mengenal pasti satu jenis *tweet* sindiran (iaitu frasa kata kerja positif diikuti oleh frasa keadaan negatif). Dan, mereka tidak berjaya mengenal pasti sindiran apabila frasa keadaan negatif diikuti oleh sentimen positif dalam ayat yang berasingan. Selain itu, keamatan negatif tidak dianggap dalam kerja mereka. Di samping itu, kerja-kerja itu tidak mempertimbangkan *hashtag* dan analisis sentimen *hashtag*. *Hashtag* adalah satu topik atau kata kunci yang ditandakan dengan *tweet*. Oleh kerana banyak *hashtag* mengandungi kekutuban, mengesan sindiran di peringkat *hashtag* akan mempunyai kesan positif ke atas klasifikasi kekutuban.

Kerja penyelidikan yang seterusnya yang dilanjutkan dalam tesis ini berfungsi berdasarkan analisis sentimen bagi *hashtag*. Penyelidik mengenal pasti *tweet* sindiran berdasarkan petunjuk sindiran dan kontra antara orientasi sentimen daripada *tweet* dan *hashtag*. Walaupun, kerja itu adalah kerja utama pada tahap analisis sentimen *hashtag*, mereka tidak menggunakan pendekatan sistemik untuk mengenalpasti petunjuk sindiran. Selain itu, penyelidikan mereka hanya berdasarkan perbezaan antara orientasi sentimen daripada *tweet* dan *hashtag*. Oleh sebab ujaran sindiran mengandungi hiperbola dan penokoktambahan dan beberapa *hashtag* digunakan untuk menekankan teks, maka, mengenal pasti berdasarkan perbezaan antara sentimen *tweet* dan *hashtag* tidak mencukupi.

Bagi menangani masalah, Model Pengesanan Sindiran (SDM) dicadangkan. Dalam model yang dicadangkan, tiga pengelas iaitu Pengelas Sindiran SentiStrength (SSC), Pengelas Hashtags Sindiran (SHC) dan Pengelas Sindiran Hashtags-SentiStrength (HSSC) digunakan. SSC berfungsi pada tahap analisis sentimen bukan *hashtag*, manakala SHC dan HSSC pada tahap analisis sentimen *hashtag*. Dalam SSC, sindiran yang telah ditetapkan mengikut tahap kekuatan *tweet*. Beberapa ciri leksikal dan pragmatik seperti emotikon, seru, kata-kata ibu dan kata-kata yang panjang digunakan dalam formula *SentiStrength* yang dicadangkan.

Pengelas Hashtags Sindiran (SHC) digunakan untuk mengenal pasti *tweet* sindiran berdasarkan Penunjuk Hashtags Sindiran (SHI) dan Analisis Sentimen Hashtags (SHA). Dalam pengelas (SHC), satu algoritma *bootstrapping* digunakan untuk mengenal pasti Penunjuk Hashtags Sindiran (SHI). SHI mengandungi senarai *hashtag* yang membantu untuk mengenal pasti *tweet* sindiran dengan mudah. Dalam model yang dicadangkan (SDM), jika *tweet* mengandungi SHI, ia akan dilabelkan sebagai *tweet* sindiran; jika tidak Analisis Sentimen Hashtags (SHA) akan digunakan. SHA bekerja berdasarkan perbezaan antara orientasi sentimen *tweet* dan *hashtag*. Dalam bahagian ini, *hashtag* ditoken melalui pra pemprosesan dan orientasi hashtag dikenalpasti. Seterusnya, orientasi *tweet* tanpa *hashtag* juga dikenalpasti. *Tweet* tersebut dianggap sebagai *hashtag* sindiran jika terdapat satu perbezaan antara orientasi *tweet* dan hashtag.

HSSC berfungsi berdasarkan tahap kekuatan *tweet* dan *hashtag*. Dalam pengelas ini, kesan sentimen hashtag untuk meningkatkan kekutuban tweet dipertimbangkan.

Model Pengesanan Sindiran (SDM) telah diuji pada dua set data yang mana setiap set data mengandungi 3000 *tweet* sindiran dan bukan sindiran. Semua *tweet* dipetik secara rawak menggunakan API Twitter. Setakat ini, tiada kerja yang dilakukan dalam pengesanan sindiran pada tahap *hashtag* dan analisis sentimen berasaskan bukan *hashtag*. Jadi, sesuatu yang baru berkenaan model yang dicadangkan (SDM) adalah dalam mengenal pasti *tweet* sindiran dengan menganalisis kekuatan *tweet* pada tahap hashtag dan analisis sentimen bukan *hashtag*. Hasil kajian (0.85% ketepatan) menunjukkan bahawa SDM adalah lebih tepat dan berkesan daripada kerja-kerja yang sedia ada yang telah dilakukan berdasarkan perbezaan antara frasa keadaan positif dan negatif analisis sentimen berasaskan *hashtag*.

ACKNOWLEDGEMENTS

Undertaking this PhD has been a truly life-changing experience for me and it would not have been possible to do without the support and guidance that I received from many people.

First and foremost, I would like to express my sincere gratitude to my supervisor Dr. Masrah Azrifah Azmi Murad for the continuous support of my study and research, for her patience, motivation, enthusiasm, and immense knowledge. Her guidance helped me in all the time of research and writing of this thesis. Her encouragement and help made me feel confident to overcome every difficulty I encountered in all the stages of this research. What I really learned from her, however, is his attitude to work and life - always aiming for excellence.

I would like to extend my gratitude and thanks to the distinguished committee member, Dr Nurfadhlina Mohd Sharef, for her advices and insightful comments, which guides me through the proper direction. I am indebted for her knowledge and helpful contributions on this thesis.

I am very grateful to the Faculty of Computer Science and Information Technology and the staff of Postgraduate office, School of Graduate Studies, Library and Universiti Putra Malaysia, for providing me excellent research environment. Thanks to every person who has supported me to pursue and finish my Ph.D.

I am very grateful to my family, my parents , my mother, my brothers, and my sister for their unflagging love and support throughout my life. I have no suitable words that can fully describe my everlasting love to them except, I love you all.

Words fail me to express my appreciation to my lovely parents whose dedication, love and persistent confidence in me, has taken the load off my shoulder. I owe them for being unselfishly let her intelligence , passions , and ambitions collide with mine.

Last but not least, it gives me immense pleasure to express my deepest gratitude to my friends, colleagues and lab mates for their unlimited support and encouragement.

Finally, I would like to thank everybody who was important to the successful realization of this thesis, as well as I express my apology that I could not mention you all personally.

I certify that a Thesis Examination Committee has met on 19 July 2016 to conduct the final examination of Samaneh Nadali on her thesis entitled "Sarcasm Detection Model Based on Tweets' Strength using Hashtag and Non-Hashtag Sentiment Analysis" in accordance with the Universities and University Colleges Act 1971 and the Constitution of the Universiti Putra Malaysia [P.U.(A) 106] 15 March 1998. The Committee recommends that the student be awarded the Doctor of Philosophy.

Members of the Thesis Examination Committee were as follows:

Rusli bin Hj Abdullah, PhD
Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Chairman)

Rodziah binti Atan, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Internal Examiner)

Norwati Mustapha, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Internal Examiner)

Hasan Davulcu, PhD
Associate Professor
Arizona State University
United States
(External Examiner)

ZULKARNAIN ZAINAL, PhD
Professor and Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 23 August 2016

This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfilment of the requirement for the degree of Doctor of Philosophy.

The members of the Supervisory Committee were as follows:

Masrah Azrifah Azmi Murad, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Chairman)

Nurfadhlina Binti Mohd Sharef, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Member)

Aida Mustafa, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Member)

____
BUJANG KIM HUAT, PhD
Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:

Declaration by graduate student

I hereby confirm that:

- this thesis is my original work;
- quotations, illustrations and citations have been duly referenced;
- this thesis has not been submitted previously or concurrently for any other degree at any other institutions;
- intellectual property from the thesis and copyright of thesis are fully-owned by Universiti Putra Malaysia, as according to the Universiti Putra Malaysia (Research) Rules 2012;
- written permission must be obtained from supervisor and the office of Deputy Vice-Chancellor (Research and Innovation) before thesis is published (in the form of written, printed or in electronic form) including books, journals, modules, proceedings, popular writings, seminar papers, manuscripts, posters, reports, lecture notes, learning modules or any other materials as stated in the Universiti Putra Malaysia (Research) Rules 2012;
- there is no plagiarism or data falsification/fabrication in the thesis, and scholarly integrity is upheld as according to the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) and the Universiti Putra Malaysia (Research) Rules 2012. The thesis has undergone plagiarism detection software.

Signature: ____________________ Date: __________

Name and Matric No.: Samaneh Nadali, GS33526

Declaration by Members of Supervisory Committee

This is to confirm that:

- the research conducted and the writing of this thesis was under our supervision
- supervision responsibilities as stated in the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) are adhered to.

Signature: ______________________

Name of Chairman of Supervisory Committee: Masrah Azrifah Azmi Murad

ASSOC. PROF. DR. MASRAH AZRIFAH AZMI MURAD
Deputy Dean
(Research and Graduate Studies)
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
43400 UPM, Serdang, Selangor

Signature: ______________________

Name of Member of Supervisory Committee: ______________________

Signature: ______________________

Name of Member of Supervisory Committee: Aida Mustapha

ASSOC. PROF. DR. MASRAH AZRIFAH AZMI MURAD
Deputy Dean
(Research and Graduate Studies)
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
43400 UPM, Serdang, Selangor

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Background

Twitter is one of the most popular platforms of microblogs which has been used for all ordinary individuals, politics and companies (Himelboim et al., 2013). Twitter allows registered users to read and post tweets (140 characters messages). All of the studies on SA of Twitter messages are term based (Go et al., 2009; Bermingham and Smeaton, 2010; Pak and Paroubek, 2010; Barbosa and Feng, 2010). Previous researchers extracted tweets based on a certain term and then analyze the sentiment of these extracted Twitter posts. In the general area of SA, sarcasm plays a role as an interfering factor that can flip the polarity of a message (Liebrecht et al., 2013).

In the Oxford English Dictionary (OED) (1989) "sarcasm" is defined as " a sharp, bitter, or cutting expression or remark; a bitter gibe or taunt". Sarcasm may employ ambivalence, although it is not necessarily ironic. Sarcasm might be used to comic effect or can be used to hurt or offend. Unlike simple negation words, a sarcasm message usually expresses a negative sentiment utilizing only positive words or even strong positive words. Although detection of sarcasm is not crucial, it is important for the development of sentiment analysis system (Liebrecht et al., 2013).

In this thesis, a sarcasm detection model for tweets was introduced. Twitter is chosen in this study, because it is one of the largest platforms where people tend to explain their opinion. Twitter also provides features such as hashtag which aid in detecting sarcasm in the tweets.

Different studies have been done in sarcasm detection such as; semi-supervised sarcasm recognition, investigation of the impact of lexical and pragmatic factors, identification of sarcasm based on intensifiers and exclamation, contrast between positive and negative situation verb phrases, identifying the relationship between a tweet and an author's past tweet, and identifying extra-linguistic information from the context of an utterance on Twitter such as properties of the author, the audience and the immediate communicative environment (Davidov et al., 2010; González-Ibánez et al., 2011; Reyes et al., 2012; Liebrecht et al., 2013; Rajadesingan et al., 2015; Bamman and Smith, 2015).

Due to the intentional ambiguity, analysis of sarcasm is a difficult task not only for machine, but even for human. Although, sarcasm detection has an important role on SA, it is generally disregard in social media analysis, because sarcasm analysis is very complicated. Since the aim of the SA is to automatically identify the sentiment of a document, misinterpreting sarcasm indicates a big challenge (Davidov et al., 2010).

1.2 Problem Statement

Lack of study in determining sarcasm among social media is one of the propounded problems from early days. Since recognizing sarcasm is important for development of sentiment analysis systems, identifying sarcastic tweets becomes an issue in this area of research. Previous studies in sarcasm detection are divided into two tasks: context-level and content-level. Several approaches have worked on the former task (Rajadesingan et al., 2015; Bamman and Smith, 2015).

The past works in sarcasm detection for latter task (content-level) involves rule-based and statistical approaches. Some studies worked based on the unigrams, pragmatic features (such as emoticon) (González-Ibánez et al., 2011; Carvalho et al., 2009; Barbieri and Saggion, 2014) whereas, other studies have worked based on extraction of common patterns, such as hashtag-based sentiment (Maynard and Greenwood, 2014), positive verb being followed by negative situation phrases (Riloff et al., 2013), or discriminative n-grams (Tsur et al., 2010; Davidov et al., 2010).

Riloff et al. (2013) used a well-constructed lexicon based method for identifying sarcasm at the level of content based on the contrast between positive sentiment and negative situation phrases (Rajadesingan et al., 2015). They presented a novel bootstrapping algorithm that automatically learns a list of positive sentiment phrases and negative situation phrases from sarcastic tweets. Their method was efficient, however, they achieved 0.62 of precision. Because they just focused on one type of tweets, i.e. positive verb phrases followed by negative situation phrases (e.g. I love being ignored). This method has some disadvantages.

Firstly, it was not able to identify other types of sarcasm; i.e. negative situation phrases followed by positive sentiment in separate sentences or clue. Secondly, the intensity of the negativity of the tweets is not identified which may be useful for identifying sarcastic tweets.

Finally this method focused only on the sentiment analysis of the tweets. Identifying sarcastic tweets based on the sentiment analysis of the hashtags is not possible. Hashtags are un-spaced phrases or words that are followed by "#". Twitter users use hashtags for using their feelings, so most of the hashtags contains polarity such as "#love", "hate" and "amazing" which can flip or enhance the polarity of the tweets.

Sarcasm detection based on the sentiment analysis of the hashtags, just worked by (Maynard and Greenwood, 2014). Their method worked based on the sarcasm indicators and contrast between the sentiment orientation of the tweets and hashtags. Although, their method was the primary study on the hashtags sentiment analysis, it is limited to identify sarcasm indicators; hashtags tokenization and finding sentiment orientation of the tweets and hashtags. Therefore, they obtained 0.46 pf precision. Only 77 sarcasm indicators were extracted by their method which is very small. Furthermore, tokenizing of some hashtags such as "#gratstart" is not possible in their method. In addition, for identifying sarcastic tweets and hashtags, lexicon based approach were used which is not accurate approach. For some event such as "going to the dentist" or "waiting for a long time" which have negative sentiment, their method was not able identify the orientations correctly. Moreover, their method is not able to identify sarcasm when there is no sarcasm indicators and no contrast between the sentiment orientation of the tweets and hashtags. Generally, people use intensifiers in their messages in order to make the expression hyperbolic and thereby sarcastic, without using a linguistic marker such as: "#sarcasm". As far as we are concerned, there is no work done in identifying English sarcastic tweets using hyperbole.

Although most of the previous works in sarcasm detection have been done by the psychologists and behavioral scientists (Gibbs and Colston, 2007; Gibbs, 1986; Kreuz and Caucci, 2007; Kreuz and Glucksberg, 1989; Utsumi, 2000), only few works have been done by social media analysis (González-Ibánez et al., 2011; Tsur et al., 2010) because sarcasm detection is a complex task. Analysis of sarcasm is usually ignored in social media analysis due to the complexity and intentional ambiguity in sarcasm. This thesis addresses the following problems:

- There is no work done on modeling sarcastic tweets detection that support hashtags-based sentiment and non-hashtags based sentiment.
- There is no work done on identifying sarcastic tweets using the strength level of the tweets at the level of the non-hashtags sentiment analysis.
- Lack of studies on determining sarcastic tweets using sarcasm indicator and contrast between the orientation of the tweets and hashtag(s) (hashtags-based sentiment analysis).
- To date, the model to detect sarcasm at the level of hashtags and non-hashtags sentiment analysis with high precision is not provided.
- lack of sarcasm indicator for identifying sarcastic tweets at the level of hashtahs sentiment analysis.
- At the level of hashtags sentiment analysis, there is no work done on identifying sarcastic tweets using the strength level of tweets and hashtags.

1.3 Research Objectives

The main objective of the research is to identify sarcastic tweets based on the strength level of the tweets. This can be done by achieving the following objectives:

- To propose a new Sarcasm Detection Model (SDM) in order to classify tweets into sarcasm and non-sarcasm at the level of hashtag and non-hashtag based sentiment analysis.
- Tp propose a new classifier named: SentiStrength Sarcasm Classifier (SSC) to classify sarcastic tweets based on the strength level of the tweets only at the level of non-hashtags sentiment analysis.
- To develop a classifier named: Sarcasm Hashtags Classifier (SHC) to classify tweets into sarcastic and non-sarcastic (at the level of the hashtags-based sentiment) using the sarcasm hashtags indicators sentiment hashtags analysis.
- To propose a model for identifying sarcastic tweets at the level of hashtags and non-hashtags sentiment analysis with high precision.
- To create more Sarcasm indicators for identifying sarcastic tweet more accurate.
- To propose a new classifier named: Hashtags- SentiStrength Sarcasm Classifier (HSSC) to classify tweets into sarcastic and non-sarcastic (at the level of the hashtags) based on the strength level of the tweets and hashtag(s).

In this study, three classifiers named: SSC, SHC and HSSC are used. These classifiers work better when they are used as part of a coherent model rather than used individually.

1.4 Research Scope

Since sarcasm detection has a positive effect on sentiment analysis (sarcasm can flip the polarity of a sentence) this research is focused on sarcasm detection on Twitter posts. Sarcastic tweets detection has been done at two different aspects:

- content-based aspect (Riloff et al., 2013)
- contextual-based aspect (Rajadesingan et al., 2015; Bamman and Smith, 2015)

In this thesis we are focusing on the content of the tweets only.

Moreover, in this study we focus only on one type of social media network. i.e Twitter.

In order to compare the proposed model (SDM), in this study, precision, recall and F-score are used (same as previous works).

1.5 Research Contribution

The contributions of this research are as follows:

- New model for sarcasm detection based on tweets' strength at the level of the hashtag(s) and non-hashtag(s).
- A classifier that can classify sarcastic tweet based on the strength level of the hashtag(s) and tweet.
- The combination of the lexical and pragmatic features to recognize the strength levels of tweet. The combination of both features has a positive effect in identifying the strength level of tweet.
- A large number of sarcasm Hashtags Indicator (SHI) is created to identify sarcastic tweets more effectively.
- A classifier that can classify sarcastic tweet based on the contrast between the orientation of the tweet and hashtags.

1.6 Overview of Thesis

This thesis is outlined in six chapters. Chapter 1 provides background information about sentiment classification and sarcasm detection approaches, and the problem statement is discussed. The objectives and contributions of this research are also included in this chapter.

Chapter 2 consists of a literature review on sentiment analysis (SA) as well as SA of microblogs, the type of existing approaches that have been presented in this area and the sarcasm detection approach.

Chapter 3 presents the research methodology of this study. The proposed Sarcasm Detection Model (SDM) is included. Evaluation metrics is discussed in this chapter.

Chapter 4 describes the proposed Sarcasm Detection Model (SDM) and the contribution of this study. This chapter illustrates how the proposed model works. Furthermore, implementation of the SDM is discussed.

Chapter 5 reports the results and discussion. Finally, Chapter 6 is the conclusion which summarizes the most important aspects of the research. This chapter ends with suggested future research.

1.7 Summary

The central goal of this thesis is to develop a new sarcasm detection model for identifying sarcastic tweets at the level of hashtag(s) and non-hashtag(s). This chapter briefly covers different approaches in sarcasm detection. Then the problem statement is explained. After that, the research objectives, contributions and scope of this research are elaborated. Finally, the chapter ends by discussing the thesis outline.

REFERENCES

Agarwal, A., Xie, B., Vovsha, I., Rambow, O., and Passonneau, R. (2011). Sentiment analysis of twitter data. In *Proceedings of the Workshop on Languages in Social Media*, pp. 30–38. Association for Computational Linguistics.

Aisopos, F., Papadakis, G., and Varvarigou, T. (2011). Sentiment analysis of social media content using n-gram graphs. In *Proceedings of the 3rd ACM SIGMM international workshop on Social media*, pp. 9–14. ACM.

Aoki, S. and Uchida, O. (2011). A method for automatically generating the emotional vectors of emoticons using weblog articles. In *Proc. 10th WSEAS Int. Conf. on Applied Computer and Applied Computational Science, Stevens Point, Wisconsin, USA*, pp. 132–136.

Attardo, S. (2000). Irony as relevant inappropriateness. *Journal of pragmatics*, 32(6):793–826.

Attardo, S., Eisterhold, J., Hay, J., and Poggi, I. (2003). Multimodal markers of irony and sarcasm. *Humor*, 16(2):243–260.

Baccianella, S., Esuli, A., and Sebastiani, F. (2010). Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *LREC*, volume 10, pp. 2200–2204.

Baeza-Yates, R. A. and Rello, L. (2011). How bad do you spell?: The lexical quality of social media. In *The Future of the Social Web*.

Bakker, S. (2015). The role of pragmatic markers in the recognition of sarcasm in tweets.

Bamman, D. and Smith, N. A. (2015). Contextualized sarcasm detection on twitter. In *Ninth International AAAI Conference on Web and Social Media*.

Barbieri, F. and Saggion, H. (2014). Modelling irony in twitter: Feature analysis and evaluation. In *Proceedings of Language Resources and Evaluation Conference (LREC)*, pp. 4258–4264.

Barbosa, L. and Feng, J. (2010). Robust sentiment detection on twitter from biased and noisy data. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, pp. 36–44. Association for Computational Linguistics.

Bekkerman, R., Bilenko, M., and Langford, J. (2011). *Scaling up machine learning: Parallel and distributed approaches*. Cambridge University Press.

Bermingham, A. and Smeaton, A. F. (2010). Classifying sentiment in microblogs: is brevity an advantage? In *Proceedings of the 19th ACM international conference on Information and knowledge management*, pp. 1833–1836. ACM.

Bifet, A. and Frank, E. (2010). Sentiment knowledge discovery in twitter streaming data. In *Discovery Science*, pp. 1–15. Springer.

Bowers, J. W. (1964). Some correlates of language intensity. *Quarterly Journal of Speech*, 50(4):415–420.

Bradley, M. M. and Lang, P. J. (1999). Affective norms for english words (anew): Instruction manual and affective ratings. Technical report, Technical Report C-1, The Center for Research in Psychophysiology, University of Florida.

Brown, P. F., Desouza, P. V., Mercer, R. L., Pietra, V. J. D., and Lai, J. C. (1992). Class-based n-gram models of natural language. *Computational linguistics*, 18(4):467–479.

Bryant, G. A. and Tree, J. E. F. (2005). Is there an ironic tone of voice? *Language and speech*, 48(3):257–277.

Bühler, K. (1990). *Theory of language: The representational function of language*, volume 25. John Benjamins Publishing.

Burgers, C. F. (2010). *Verbal irony: Use and effects in written discourse*. [Sl: sn].

Capelli, C. A., Nakagawa, N., and Madden, C. M. (1990). How children understand sarcasm: The role of context and intonation. *Child Development*, 61(6):1824–1841.

Carvalho, P., Sarmento, L., Silva, M. J., and De Oliveira, E. (2009). Clues for detecting irony in user-generated contents: oh...!! it's so easy;-). In *Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion*, pp. 53–56. ACM.

Chang, C.-C. and Lin, C.-J. (2011). Libsvm: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3):27.

Clark, H. H. (1996). Using language. 1996. *Cambridge University Press: Cambridge)*, 952:274–296.

Cohen, J. et al. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1):37–46.

Cui, A., Zhang, M., Liu, Y., and Ma, S. (2011). Emotion tokens: Bridging the gap among multilingual twitter sentiment analysis. In *Information retrieval technology*, pp. 238–249. Springer.

Cunningham, H., Maynard, D., Bontcheva, K., and Tablan, V. (2002). Gate: an architecture for development of robust hlt applications. In *Proceedings of the 40th annual meeting on association for computational linguistics*, pp. 168–175. Association for Computational Linguistics.

Davidov, D., Tsur, O., and Rappoport, A. (2010). Enhanced sentiment learning using twitter hashtags and smileys. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, pp. 241–249. Association for Computational Linguistics.

Ding, X., Liu, B., and Yu, P. S. (2008). A holistic lexicon-based approach to opinion mining. In *Proceedings of the 2008 International Conference on Web Search and Data Mining*, pp. 231–240. ACM.

Dresner, E. and Herring, S. C. (2010a). Functions of the nonverbal in cmc: Emoticons and illocutionary force. *Communication theory*, 20(3):249–268.

Dresner, E. and Herring, S. C. (2010b). Functions of the nonverbal in cmc: Emoticons and illocutionary force. *Communication theory*, 20(3):249–268.

Esuli, A. and Sebastiani, F. (2006a). Determining term subjectivity and term orientation for opinion mining. In *EACL*, volume 6, p. 2006.

Esuli, A. and Sebastiani, F. (2006b). Sentiwordnet: A publicly available lexical resource for opinion mining. In *Proceedings of LREC*, volume 6, pp. 417–422. Citeseer.

Gibbs, R. and Izett, C. (2005). Irony as persuasive communication. *Figurative language comprehension: Social and cultural influences*, pp. 131–151.

Gibbs, R. W. (1986). On the psycholinguistics of sarcasm. *Journal of Experimental Psychology: General*, 115(1):3.

Gibbs, R. W. and Colston, H. L. (2007). *Irony in language and thought: A cognitive science reader*. Psychology Press.

Giora, R. (1995). On irony and negation. *Discourse processes*, 19(2):239–264.

Go, A., Bhayani, R., and Huang, L. (2009). Twitter sentiment classification using distant supervision. *CS224N Project Report, Stanford*, 1:12.

González-Ibánez, R., Muresan, S., and Wacholder, N. (2011). Identifying sarcasm in twitter: a closer look. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*, pp. 581–586. Association for Computational Linguistics.

Grice, H. (1975). Logic and conversation'in p. cole and j. morgan (eds.) syntax and semantics volume 3: Speech acts.

Grice, H. P. (1970). *Logic and conversation*. na.

Grice, H. P. (1978). Some further notes on logic and conversation.

Günther, T. and Furrer, L. (2013). Gu-mlt-lt: Sentiment analysis of short messages using linguistic features and stochastic gradient descent.

Guthrie, D., Allison, B., Liu, W., Guthrie, L., and Wilks, Y. (2006). A closer look at skip-gram modelling. In *Proceedings of the 5th international Conference on Language Resources and Evaluation (LREC-2006)*, pp. 1–4.

Hamamoto, H. (1998). Irony from a cognitive perspective. *PRAGMATICS AND BEYOND NEW SERIES*, pp. 257–270.

Hancock, J. T. (2004). Verbal irony use in face-to-face and computer-mediated conversations. *Journal of Language and Social Psychology*, 23(4):447–463.

Hansen, L. K., Arvidsson, A., Nielsen, F. Å., Colleoni, E., and Etter, M. (2011). Good friends, bad news-affect and virality in twitter. In *Future information technology*, pp. 34–43. Springer.

Himelboim, I., McCreery, S., and Smith, M. (2013). Birds of a feather tweet together: Integrating network and content analyses to examine cross-ideology exposure on twitter. *Journal of Computer-Mediated Communication*, 18(2):40–60.

Huddleston, R., Pullum, G. K., et al. (2002). The cambridge grammar of english. *Language. Cambridge: Cambridge University Press*, pp. 1–23.

Ishii, N., Tsuchiya, E., Bao, Y., and Yamaguchi, N. (2005). Combining classification improvements by ensemble processing. In *Software Engineering Research, Management and Applications, 2005. Third ACIS International Conference on*, pp. 240–246. IEEE.

Jahandarie, K. (1999). *Spoken and written discourse: A multi-disciplinary perspective*. Number 1. Greenwood Publishing Group.

Kotthoff, H. (2003). Responding to irony in different contexts: On cognition in conversation. *Journal of pragmatics*, 35(9):1387–1411.

Kreuz, R. J. (1996). The use of verbal irony: Cues and constraints. *Metaphor: Implications and applications*, pp. 23–38.

Kreuz, R. J. and Caucci, G. M. (2007). Lexical influences on the perception of sarcasm. In *Proceedings of the Workshop on computational approaches to Figurative Language*, pp. 1–4. Association for Computational Linguistics.

Kreuz, R. J. and Glucksberg, S. (1989). How to be sarcastic: The echoic reminder theory of verbal irony. *Journal of Experimental Psychology: General*, 118(4):374.

Kreuz, R. J. and Link, K. E. (2002). Asymmetries in the use of verbal irony. *Journal of Language and Social Psychology*, 21(2):127–143.

Kumar, A. and Sebastian, T. M. (2012). Sentiment analysis on twitter. *IJCSI International Journal of Computer Science Issues*, 9(3):372–378.

Liebrecht, C., Kunneman, F., and van den Bosch, A. (2013). The perfect solution for detecting sarcasm in tweets# not.

Liu, B., Hu, M., and Cheng, J. (2005). Opinion observer: analyzing and comparing opinions on the web. In *Proceedings of the 14th international conference on World Wide Web*, pp. 342–351. ACM.

Liu, B. and Zhang, L. (2012). A survey of opinion mining and sentiment analysis. In *Mining text data*, pp. 415–463. Springer.

Liu, K.-L., Li, W.-J., and Guo, M. (2012). Emoticon smoothed language models for twitter sentiment analysis. In *AAAI*.

Livne, A., Simmons, M. P., Adar, E., and Adamic, L. A. (2011). The party is over here: Structure and content in the 2010 election. *ICWSM*, 11:17–21.

Lukin, S. and Walker, M. (2013). Really? well. apparently bootstrapping improves the performance of sarcasm and nastiness classifiers for online dialogue. In *Proceedings of the Workshop on Language Analysis in Social Media*, pp. 30–40.

Marchetti-Bowick, M. and Chambers, N. (2012). Learning for microblogs with distant supervision: Political forecasting with twitter. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 603–612. Association for Computational Linguistics.

Maynard, D., Bontcheva, K., and Rout, D. (2012). Challenges in developing opinion mining tools for social media. *Proceedings of the@ NLP can u tag# usergeneratedcontent*, pp. 15–22.

Maynard, D. and Greenwood, M. A. (2014). Who cares about sarcastic tweets? investigating the impact of sarcasm on sentiment analysis. In *Proceedings of LREC*.

Miller, G. and Fellbaum, C. (1998). Wordnet: An electronic lexical database.

Muecke, D. C. (1978). Irony markers. *Poetics*, 7(4):363–375.

Nielsen, F. Å. (2011). A new anew: Evaluation of a word list for sentiment analysis in microblogs. *arXiv preprint arXiv:1103.2903*.

O'Connor, B., Balasubramanyan, R., Routledge, B. R., and Smith, N. A. (2010). From tweets to polls: Linking text sentiment to public opinion time series. *ICWSM*, 11(122-129):1–2.

Owoputi, O., O'Connor, B., Dyer, C., Gimpel, K., Schneider, N., and Smith, N. A. (2013). Improved part-of-speech tagging for online conversational text with word clusters. Association for Computational Linguistics.

Pak, A. and Paroubek, P. (2010). Twitter as a corpus for sentiment analysis and opinion mining. In *LREC*, volume 10, pp. 1320–1326.

Partington, A. (2007). Irony and reversal of evaluation. *Journal of Pragmatics*, 39(9):1547–1569.

Pennebaker, J. W., Booth, R. J., and Francis, M. E. (2007). Linguistic inquiry and word count: Liwc [computer software]. *Austin, TX: liwc. net*.

Petrovic, S., Osborne, M., and Lavrenko, V. (2010). The edinburgh twitter corpus. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Linguistics in a World of Social Media*, pp. 25–26.

Qadir, A. and Riloff, E. (2014). Learning emotion indicators from tweets: Hashtags, hashtag patterns, and phrases. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), Association for Computational Linguistics*, pp. 1203–1209.

Rajadesingan, A., Zafarani, R., and Liu, H. (2015). Sarcasm detection on twitter: A behavioral modeling approach. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, pp. 97–106. ACM.

Rakov, R. and Rosenberg, A. (2013). " sure, i did the right thing": a system for sarcasm detection in speech. In *INTERSPEECH*, pp. 842–846.

Renkema, J. (2004). *Introduction to discourse studies*. John Benjamins Publishing.

Reyes, A., Rosso, P., and Buscaldi, D. (2012). From humor recognition to irony detection: The figurative language of social media. *Data & Knowledge Engineering*, 74:1–12.

Reyes, A., Rosso, P., and Veale, T. (2013). A multidimensional approach for detecting irony in twitter. *Language Resources and Evaluation*, 47(1):239–268.

Riloff, E., Qadir, A., Surve, P., De Silva, L., Gilbert, N., and Huang, R. (2013). Sarcasm as contrast between a positive sentiment and negative situation. In *EMNLP*, pp. 704–714.

Rockwell, P. (2003). Empathy and the expression and recognition of sarcasm by close relations or strangers. *Perceptual and motor skills*, 97(1):251–256.

Rockwell, P. (2007). Vocal features of conversational sarcasm: A comparison of methods. *Journal of psycholinguistic research*, 36(5):361–369.

Saif, H., He, Y., and Alani, H. (2011). Semantic smoothing for twitter sentiment analysis.

Saif, H., He, Y., and Alani, H. (2012a). Alleviating data sparsity for twitter sentiment analysis. CEUR Workshop Proceedings (CEUR-WS. org).

Saif, H., He, Y., and Alani, H. (2012b). Semantic sentiment analysis of twitter. In *The Semantic Web–ISWC 2012*, pp. 508–524. Springer.

Seto, K.-i. (1998). On non-echoic irony. *PRAGMATICS AND BEYOND NEW SERIES*, pp. 239–256.

Shamma, D. A., Kennedy, L., and Churchill, E. F. (2009). Tweet the debates: understanding community annotation of uncollected sources. In *Proceedings of the first SIGMM workshop on Social media*, pp. 3–10. ACM.

Sperber, D. and Wilson (1986). *Relevance: Communication and cognition*, volume 1. Citeseer.

Speriosu, M., Sudan, N., Upadhyay, S., and Baldridge, J. (2011). Twitter polarity classification with label propagation over lexical links and the follower graph. In *Proceedings of the First workshop on Unsupervised Learning in NLP*, pp. 53–63. Association for Computational Linguistics.

Strapparava, C., Valitutti, A., et al. (2004). Wordnet affect: an affective extension of wordnet. In *LREC*, volume 4, pp. 1083–1086.

Takala, P., ekka Malo, P., Sinha, A., and Ahlgren, O. (2014). Gold-standard for topic-specific sentiment analysis of economic text s. In *LREC*, volume 2014, pp. 2152–2157.

Thelwall, M. (2013). Heart and soul: Sentiment strength detection in the social web with sentistrength.(2013).

Thelwall, M., Buckley, K., and Paltoglou, G. (2012a). Sentiment strength detection for the social web. *Journal of the American Society for Information Science and Technology*, 63(1):163–173.

Thelwall, M., Buckley, K., and Paltoglou, G. (2012b). Sentiment strength detection for the social web. *Journal of the American Society for Information Science and Technology*, 63(1):163–173.

Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., and Kappas, A. (2010). Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*, 61(12):2544–2558.

Tsur, O., Davidov, D., and Rappoport, A. (2010). Icwsm-a great catchy name: Semi-supervised recognition of sarcastic sentences in online product reviews. In *ICWSM*.

Utsumi, A. (2000). Verbal irony as implicit display of ironic environment: Distinguishing ironic utterances from nonirony. *Journal of Pragmatics*, 32(12):1777–1806.

Vapnik, V. (1998). Statistical learning theory. 1998.

Wallace, B. C., Do Kook Choe, L. K., and Charniak, E. (2014). Humans require context to infer ironic intent (so computers probably do, too). In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 512–516.

Wang, X., Wei, F., Liu, X., Zhou, M., and Zhang, M. (2011). Topic sentiment analysis in twitter: a graph-based hashtag sentiment classification approach. In *Proceedings of the 20th ACM international conference on Information and knowledge management*, pp. 1031–1040. ACM.

Whalen, J. M., Pexman, P. M., and Gill, A. J. (2009). âĂIJshould be funâĂŤnot!âĂİ: Incidence and marking of nonliteral language in e-mail. *Journal of Language and Social Psychology*.

Wharton, T. (2003). Interjections, language, and the âĂŸshowing/sayingâĂŹcontinuum. *Pragmatics & Cognition*, 11(1):39–91.

Wiebe, J., Wilson, T., and Cardie, C. (2005). Annotating expressions of opinions and emotions in language. *Language resources and evaluation*, 39(2-3):165–210.

Witten, I. H. and Frank, E. (2005). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.

Woodland, J. and Voyer, D. (2011). Context and intonation in the perception of sarcasm. *Metaphor and Symbol*, 26(3):227–239.

Zhang, L., Ghosh, R., Dekhil, M., Hsu, M., and Liu, B. (2011). Combining lexiconbased and learning-based methods for twitter sentiment analysis. *HP Laboratories, Technical Report HPL-2011*, 89.

Zhao, J., Dong, L., Wu, J., and Xu, K. (2012). Moodlens: an emoticon-based sentiment analysis system for chinese tweets. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 1528–1531. ACM.

Zhu, X., Lafferty, J., and Rosenfeld, R. (2005). *Semi-supervised learning with graphs*. Carnegie Mellon University, language technologies institute, school of computer science.