MEASURING OPINION CREDIBILITY IN TWITTER

BY

MYA THANDAR

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE (ENGINEERING AND TECHNOLOGY)

SIRINDHORN INTERNATIONAL INSTITUTE OF TECHNOLOGY
THAMMASAT UNIVERSITY
ACADEMIC YEAR 2015

MEASURING OPINION CREDIBILITY IN TWITTER

BY

MYA THANDAR

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE (ENGINEERING AND TECHNOLOGY)

SIRINDHORN INTERNATIONAL INSTITUTE OF TECHNOLOGY

THAMMASAT UNIVERSITY

ACADEMIC YEAR 2015

MEASURING OPINION CREDIBILITY IN TWITTER

A Thesis Presented

By

MYA THANDAR

Submitted to

Sirindhorn International Institute of Technology

Thammasat University

In partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE (ENGINEERING AND TECHNOLOGY)

Approved as to style and content by

Advisor and Chairperson of Thesis Committee ______________________

(Dr.Sasiporn Usanavasin, Ph.D.)

Committee Member and
Chairperson of Examination Committee ______________________

(Asst. Prof. Pakinee Aimmanee, Ph.D.)

Committee Member ______________________

(Asst. Prof. Thepparit Banditwattanawong, Ph.D)

December 2015

Abstract

MEASURING OPINION CREDIBILTIY IN TWITTER

by

MYA THANDAR

B.C.Sc., School of Information Computer and Communication Technology (ICT), University of Computer Studies (UCSY), Yangon, Myanmar, 2009

Explosion of social media is a way of connecting people with each other and also fast growing with effects on individual behaviour. Sentiment analysis is a main point of the social media research including Twitter, Facebook, blogs, and user forums. Nowadays most of the sentiment analysing focus on opinion classification and summarization the opinions result. In sentiment analysing, most researches focus on only opinion classification and summarization of opinion results. However when we classify sentiment information, which information can identify whether this opinion is credible or not? Do we believe result of this topic can be positive if positive results come from unreliable source and negative results come from reliable source in sentiment analysing? For this reason, we propose a method for calculating of opinion credibility based on authors' expert knowledge. In our method we have two components: sentiment analyzer and opinion credibility calculator. We use two machine learning techniques: Naïve Bayes (NB) and Support Vector Machine (SVM) for sentiment analyzer. To define expert knowledge, we use author's profile, list features and author's tweets behaviour for weight factors in twitter. In opinion calculator, we produce opinion credibility value combine with sentiment analyzer result and expert knowledge score. Thereafter we converse all of our defined author's expert weight factors to analyses which weight factors weightier than others and which factors can support more credible when we compute opinion credibility. To evaluate our method we used weighted kappa statistic for accuracy of opinion credibility.

Keywords: conversion weight, expert knowledge, machine learning, opinion credibility, opinion mining, sentiment analyzing, twitter, weight factors.

Acknowledgements

I would like to gratefully and sincerely thank my adviser, Dr. Sasiporn Usanavasin for her guidance, understanding, patience, and most importantly, her invaluable constructive criticism and friendly advice during the period of study.

Also I would like to thank my committee members, Asst. Prof. Dr. Pakinee Aimmanee and Asst. Director. Dr. Thepparit Banditwattanawong for their valuable information, several supports and guidance. I would also like to thank everyone from our faculty members who supported me throughout this research. I am sincerely grateful to them for sharing their truthful and illuminating views on a number of issues related to the thesis.

I also have to thank my parents for their love and support throughout my life and give me strength to reach for the stars and chase my dreams. My sister and my family deserve my wholehearted thanks as well.

To all my friends, thank you for your understanding and encouragement in my many, many moments of crisis. Your friendship makes my life a wonderful experience. I cannot list all the names here, but you are always in my mind.

The last one is thanks to my university, Sirindhorn International Institute of Technology (SIIT), Thammasat University (TU) that gives me chance to get Master Degree. This thesis partially supported by Center of Excellence in Intelligent Informatics, Speech and Language Technology and Service Innovation (CILS) and by NRU grant at SIIT, Thammasat University.

Table of Contents

Chapter	Title	Page
	Signature Page	i
	Abstract	ii
	Acknowledgements	iii
	Table of Contents	iv
	List of Figures	vi
	List of Tables	vii
1	Introduction	1
	1.1 Credibility for Sentiment Analysis	2
2	Literature Review	4
	2.1 Sentiment Analysis	4
	2.1.1 Different levels of Sentiment Analysis	5
	2.1.2 Related Works	7
	2.2 Expert Knowledge	8
	2.2.1 Related Works for using expert knowledge	8
	2.3 Support Vector Machine (SVM)	9
	2.3.1 Linear SVM	10
	2.4 Naïve Bayes (NB)	11
	2.5 Kappa Statistic	12
	2.5.1 Weighted Kappa Statistic	12
3	Measuring Opinion Credibility in Twitter	15
	3.1 Preprocessing	15
	3.2 Sentiment Analyzer	16

3.3 Opinion Credibility Calculator 17
3.4 Conversion of Weighted Factors 19

4 Result and Discussion 21

4.1 Experiment result with HongKong Dataset 21
4.2 Experiment result with Obamacare Dataset 23

5 Conclusion 28

References 29

Appendices 31

List of Tables

Tables **Page**

2.1 Weights used for weighted kappa statistic to evaluate second-by-second agreement between two scorers of a PSG recording 12
4.1 Training Dataset 19
4.2 Result of Training Classification of HK Dataset 19
4.3 Example of trusted Author group 20
4.4 Sample Results of opinion credibility 21
4.5 Results of Training Classification for Naïve Bayes (NB) 22
4.6 Results of Training Classification for SVM 22
4.7 Levels of agreement measure for Kappa statistic (K_w) 22
4.8 Number of Agreement between system and rater for Positive result 23
4.9 Number of Agreement between system and rater for Negative result 23
4.10 Result of kappa statistic (K_w) between System and Rater for each conversion of weight factors 24

List of Figures

Figures **Page**

2.1 SVM separating hyperplanes 9
2.2 SVM maximum separating hyperplane with margin 10
3.1 Overview system of credibility result of tweet polarity 14
4.1 Example of how to measure opinion credibility 20

Chapter 1
Introduction

The world is rapidly changing with the explosion of social media nowadays. The social media has connected people together, making the world smaller and also has significantly changed people behavior. People can openly express their thoughts and opinions about many things or topics using social media applications such as Twitter, Facebook, etc. These opinions are useful and valuable for many purposes especially for businesses or organizations to improve and innovate their products or services based on positive and negative viewpoints from their customers. On the other hand, many people use social media as the way to perceive news or get better understanding of some events or topics. Being able to obtain feedbacks or opinions from people can be very beneficial to organizations and government to rethink or reform some policy or business models.

For example, 2014 Hong Kong Protest, known as the Umbrella Movement or Umbrella Revolution, began in September 2014. This revolution happened because of the China's authorities would like to limit who Hong Kongers can elect in 2017 elections. In this event, the social media is used as tools for Hong Kongers to express their opinions about this situation. They posted information, opinions and many discussions about this protest into many social media applications (e.g., Twitter, Facebook, etc.). These posts were uploaded in various forms such as texts, pictures, and videos all over the world through the Internet. An opinion poll conducted by Chinese University showed that 46% did not support this revolution, while 31% backed the civil disobedience movement[1].

Opinion mining is very important technique for getting the information about what people feel or think about some products, services, policy, news or events. It becomes one of the hot research topics in social media today. Sentiment analysis is a part of opinion mining, which is an approach for monitoring and classifying polarities of the opinions. Sentiment Analysis uses Natural Language Processing (NLP) and

[1] http://www.scmp.com/news/hong-kong/article/1597646/one-five-hongkongers-may-emigrate-over-political-reform-ruling

Information Extraction techniques to obtain writers' feeling that is expressed as positive or negative comments, questions or requests. For example,

- A huge thanks to the many protesters who were extra supportive today. Please stay safe and strong. You are all awesome #OccupyCentral
- More tourists are flocking to Hong Kong to watch the pro–democracy protesters, so business is good here. Merchants are happy.

These two tweets are positive and negative text. Sentiment analysis is used to identify whether the opinion expressed in a text is positive or negative. Thus, social media and sentient analysis are important in marketing, financial, political, public relations, health care, education, etc. Many social media tools focus on sentiment analysis to improve their knowledge about people's feelings and thoughts. These social media applications include Twitter, Facebook, blogs, and user forums.

1.1 Credibility For Sentiment Analysis

People have different opinions and perceptions about various topics, political issues, religions and events based on individual background knowledge, personal experience and beliefs. People today can post information, comments and opinions openly on Internet and some of them are reliable and some are not. For other people to perceive such information, the creditability of those posted information on social media should be considered. With today Internet, we are dealing with many potentially biased sources of information, therefore it is necessary for readers to be able to know which information is reliable and creditable in order to learn about the topics, events, or situations. For example, during 2014 Hong Kong Protest, government officers, journalists and news channels posted various information, comments and opinions about Hong Kong Revolution over social media. In that event, it is difficult to know whose opinions are more reliable or credential. One factor that can be used to identify the creditability of the posted information should be expert knowledge or reputation of the writers. To tackle this issue about creditability of online information on social media, we propose a new method to identify the credibility of comments and opinions on Twitter based on expert knowledge of the author or the writer. This creditability score is computed based on six weighted factors. Based on the creditability score, the comments or opinions about a topic can be ranked based on the highest to lowest scores.

In our method, it is composed of two components, which are sentiment analyzer and opinion credibility calculator. In the sentiment analyzer, we classify tweets sentiment polarity using two machine learning techniques: Support Vector machine (SVM) and Naïve Bayes (NB). To define expert knowledge, we use author's profile, list features and author's tweets behavior as weighted factors. In the opinion credibility calculator, we compute opinion credibility value by combining the result from the sentimental analyzer with expert knowledge score. We performed many experiments to find which factors most significantly impact on the creditability score and to evaluate our method we also use the weighted Kappa Statistic for compute the degree of agreement between human judgments and the results from our approach.

Chapter 2
Literature Review

In this chapter, we had reviewed many previous works in Sentiment Analysis, Expert Knowledge, Support Vector Machine (SVM), Naïve Bayes (NB) and Kappa Statistic. Sentiment classification is the binary classification task of labelling an opinionated document as expressing either an overall positive or an overall negative opinion. Generally speaking, sentiment analysis aims to determine the attitude of a speaker or a writer with respect to some topic or the overall contextual polarity of a document. The attitude may be his or her judgment or evaluation affective state (that is to say, the emotional state of the author when writing), or the intended emotional communication (that is to say, the emotional effect the author wishes to have on the reader). Support Vector Machine and Naïve Bayes are used for classification in machine learning and the kappa statistic is a metric that compares an observed accuracy with an expected accuracy (random chance). The kappa statistic is used not only to evaluate a single classifier, but also to evaluate classifiers amongst themselves.

2.1 Sentiment Analysis

Sentiment analysis is the task of analyzing people's opinions, sentiments, evaluations, appraisals, attitudes, and emotion also called opinion mining. They have different function such as sentiment analysis, opinion mining, opinion extraction, sentiment mining, and subjectivity analysis, affect analysis, emotion analysis and review mining, etc. In human life, they are important influencers of our behaviors. Our manner and the selection we make and our perceptions of reality rely on the decision making process. When we make a decision, we need to find out the opinions of other people (seek opinions from friends and family) or organizations (use surveys, focus groups, opinion polls, consultants). There is an enormous opinionated data in the social media on the Web. Therefore, sentiment analysis is the main point of the social media research including Twitter, Facebook, message boards, blogs, and user forums. It also effect on management science, political science, and economic and social sciences. Sentiment analysis is one of the hottest research areas in computer science.

When an organization or a business needed public or consumer opinions, it conducted surveys, opinion polls, and focus groups. Acquiring public and consumer opinions has long been a huge business itself for marketing, public relations and political campaign companies. With the explosive growth of social media (e.g., reviews, forum discussions, blogs, micro-blogs, Twitter, comments and postings in social network sites) on the Web, individuals and organizations are increasingly using the content in these media for decision marking. Nowadays, if one wants to buy a consumer product, one is no longer limited to asking one's friends and family for opinions because there are many user reviews and discussions in public forums on the Web about the product. For an organization, it may no longer be necessary to conduct surveys, opinion polls, and focus groups in order to gather public opinions because there is an abundance of such information publicly available. However, finding and monitoring opinion sites on the Web and distilling the information contained in them remains a formidable task because of the proliferation of diverse sites. Each site typically contains a huge volume of opinion text that I not always easily deciphered in long blogs and forum postings. The average human reader will have difficulty identifying relevant sites and extracting and summarizing the opinions in them. Automated sentiment analysis systems are thus needed.

2.1.1 Different Levels of Sentiment Analysis

The main research problems based on the level of the existing research; in general sentiment analysis can be examine three levels [14]: document-level sentiment analysis, sentence-level sentiment analysis and entity and aspect level.

- **Document-Level Sentiment Analysis**: The task at this level is to classify whether a whole opinion document expresses a positive or negative sentiment. For example, given a product review, the system determines whether the review expresses an overall positive or negative opinion about the product. This task is known as document-level sentiment classification. This level of analysis assumes that each document expresses opinions on a single entity (e.g., a single product). Thus, it is not applicable to documents, which evaluate or compare multiple entities.
- **Sentence-Level Sentiment Analysis:** The task at this level goes to the sentences and determines whether each sentence expressed a positive, negative, or neutral opinion. Neutral usually means no opinion. This level of analysis is closely related to subjectivity classification, which distinguishes sentences (called objective sentences) that express factual information from sentences (called subjective sentences) that express subjective views and opinions. However, we should note that subjectivity is not equivalent to sentiment as many objective sentences can imply opinions, e.g., "We bought the car last month and the windshield wiper has fallen off." Researchers have also analyzed clauses, but the clause level is still not enough, e.g., "Apple is doing very well in this lousy economy."
- **Entity and Aspect Level**: Both the document level and the sentence level analyses do not discover what exactly people liked and did not like. Aspect level performs finer-grained analysis. Aspect level was earlier called feature level (feature-based opinion mining and summarization). Instead of looking at language constructs (documents, paragraphs, sentences, clauses or phrases), aspect level directly looks at the opinion itself. It is based on the idea that an opinion consists of a sentiment (positive or negative) and a target (of opinion).An opinion without its target being identified is of limited use. Realizing the importance of opinion targets also helps us understand the sentiment analysis problem better. For example, although the sentence "although the service is not that great, I still love this restaurant" clearly has a positive tone, we cannot say that this sentence is entirely positive. In fact, the

sentence is positive about the restaurant (emphasized), but negative about its service (not emphasized). In many applications, entities and/or their different aspects describe opinion targets. Thus, the goal of this level of analysis is to discover sentiments on entities and/or their aspects. For example, the sentence "The iPhone's call quality is good, but its battery life is short" evaluates two aspects; call quality and battery life, of iPhone (entity). The sentiment on iPhone's call quality is positive, but the sentiment on its battery life is negative. The call quality and battery life of iPhone are the opinion targets. Based on this level of analysis, a structured summary of opinions about entities and their aspects can be produced, which turns unstructured text to structured data and can be used for all kinds of qualitative and quantitative analyses.

2.1.2 Related Works

Both the document level and sentence level classifications are already highly challenging. The aspect-level is even more difficult. Garcia-Moya et al., [7] address the aspect-based summarization task for retrieving product features from a collection of free text customer reviews about a product or service. Hu M. and Liu B. [6] generate feature-based summaries (FBS) of customer reviews of product sold online. They used Tree Map visualization (a method for displaying hierarchical data by using nested rectangles) to display clusters and their associated sentiment. Bafna K. and Toshniwal D. [8] deal with sentiment analysis by generating features based on summarization of customer's opinions. They used an association mining technique top fine frequent feature and used the opinion lexicon and sentiword-net to identify semantic polarities for product review. Pang, B. [9] used three machine learning techniques: Naïve Bayes, Maximum entropy and Support Vector Machine (SVM) to classify movie review documents for sentiment analysis. Pang, B., Lee, L. and Vaithyanathan S. [10] performed a linguistic analysis of the collected corpus and determine positive, negative and neutral sentiment based on the multinomial Naïve Bayes classifier that uses N-gram and POS-tags as features. Neethu M. S. and Rajasree R. [11] classified the tweets as positive, negative and extracted peoples' opinion about electronic products using Nave Bayes and SVM classifier. Liang P.-W.

and Dai, B.-R. [12] design a system called opinion miner which integrated machine learning techniques and domain-specific data. They used unigram Naïve Bayes for extracting tweets and determined that is an opinion or not. They used Mutual Information and X2 feature selection for short text classification to discard some useless features. Go A. et al., [13] classify tweets with emoticons for distant supervised learning.

2.2 Expert Knowledge

We often want the judgment of "experts" as we make important decisions in life, health, and business. But what exactly is an expert? “Expert knowledge” is what qualified individuals know as a result of their technical practices, training, and experience [15]. It may include recalled facts or evidence, inferences made by the expert on the basis of “hard facts” in response to new or undocumented situations, and integration of disparate sources in conceptual models to address system-level issues[16] . For a more detailed discussion of expert knowledge. Experts are usually identified on the basis of qualifications, training, experience, professional memberships, and peer recognition [17]. Experts provide knowledge informally when they specify information “off the top of their heads”. Informal, subjective judgments are often incorporated into scientific decisions through the selection of which problem needs to be analyzed, how the problem is to be structured, what data sources to draw upon, how results are interpreted, and what actions are recommended.

2.2.1 Related Works for Using Expert Knowledge

To explore expertise, it has little research area in social media and also it is ambiguous to decide characterize of expert. What kinds of background knowledge is impact their authority? Weng J. et al. [18] created TwitterRank, to identify influential users of Twitter. They utilized Latent Dirichlet Allocation (LDA) to calculate topical distribution of a user and made weighted user graph that point out topical similarity of users. They used PageRank algorithm to identify authorities on each topic. Pal A. and Counts S. [19] also proposed to find the most interesting and authoritative authors for any given topic in Twitter. They analyzed features such as user’s content and combined it their friends and followers information to be affective to find topical experts. They ran Gaussian Ranking Algorithm to cluster users into two clusters over their features space for finding the most authoritative users. Sharma N. K.et al. [3] designed who-is-who service to infer characterize of individual Twitter users using Twitter List Features which allow a user to make groups of other users who related on a topic. List meta-data (names and description) provided to get semantic cues about who the users in it. Further it could also infer topical expertise users analyzing the meta-data of crowdsourced Lists that contains the user. Liang C.et al [20] designed a framework to help identify misinformation with the assessments of experts. They proposed a tag-based method extracting extracted the expertise of users from their microblog contents and matched the experts with given suspected misinformation. With the judgment of experts they defined the credibility of information and confuting of misinformation. Namihira Y. et al. [21] proposed to access the credibility of information based on topic and opinion classification depending on user's knowledge (expertise). They believed if they considered tweet of user knowledge, they handled as a more reliable opinion even if it is a minor opinion. Likewise our work is also focus on a user expert knowledge but different approach. We desire to assess the credibility of tweets opinions.

2.3 Support Vector Machine (SVM)

In machine learning, support vector machine (SVM) is a computer algorithm that learns by example to assign labels to objects. It constructs a hyperplane or set of hyperplanes in a high- or infinite-dimensional space, used for classification and regression analysis. Given a set of training examples, each marked for belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other, making it a non-probabilistic binary linear classifier. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on. SVMs can also perform a non-linear classification using kernel method, implicitly mapping their inputs into high-dimensional feature spaces.

SVM is a useful and popular technique for data classification. A classification task usually involves separating data into training and testing sets. Each instance in the training set contains one target value (the class labels) and several attributes (the features or observed variables) [5]. The goal of SVM is to produce a model (based on the training data) which predicts the target values of the test data given only the test data attributes. In our research, we use linear SVM and we continue to explain for Linear SVM.

2.3.1 Linear SVM

Linear SVM is a linearly scalable routine meaning that it creates an SVM model in a CPU time which scales linearly with the size of the training data set. Suppose some given data points each belong to one of two classes, and the goal is to decide which class a new data point will be in. In the case of support vector machines, a data point is viewed as a *P*-dimensional vector (a list of *P* numbers), and we want to know whether we can separate such points with a (*P*-1)-dimensional hyperplane. This is called a linear classifier.

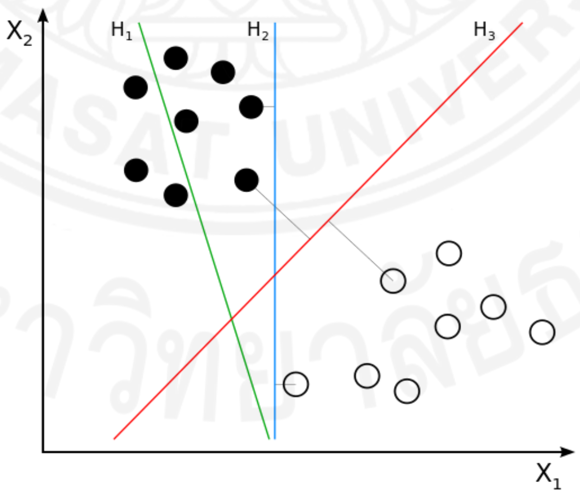

Fig 2.1: SVM separating hyperplanes[2]

There are many hyperplanes that might classify the data. One reasonable choice as the best hyperplane is the one that represents the largest separation, or

[2] https://commons.wikimedia.org/wiki/File:Svm_separating_hyperplanes.png#file

margin, between the two classes. In Figure 2.1, how a support vector machine would choose a separating hyperplane for two classes of points in 2D. H1 does not separate the classes. H2 does, but only with a small margin. H3 separates them with the maximum margin.

So we choose the hyperplane so that the distance from it to the nearest data point on each side is maximized. If such a hyperplane exists, it is known as the maximum-margin hyperplane and see in Figure in 2.2. The linear classifier, it defines is known as a maximum margin classifier; or equivalently, the perceptron of optimal stability.

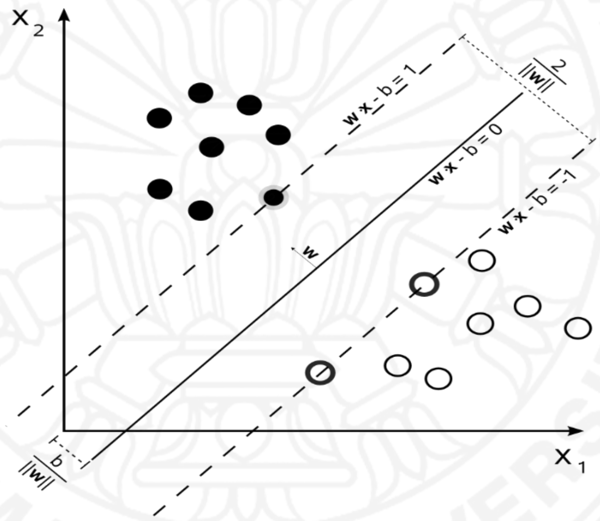

Fig 2.2: SVM maximum separating hyperplane with margin[3]

2.4 Naïve Bayes (NB)

Naive Bayes classifier is a simple probabilistic classifier which is based on Bayes theorem with strong and naïve independence assumptions. It is one of the most basic text classification techniques with various applications in email spam detection, personal email sorting, document categorization, sexually explicit content detection, language detection and sentiment detection. Despite the naïve design and oversimplified assumptions that this technique uses, Naive Bayes performs well in many complex real-world problems.

---

[3]https://en.wikipedia.org/wiki/Support_vector_machine#/media/File:Svm_max_sep_hyperplane_with_margin.png

The Naive Bayes classifier assumes that the features used in the classification are independent. Naive Bayes classifiers are highly scalable, requiring a number of parameters linear in the number of variables (features/predictors) in a learning problem. Bayes theorem provides a way of calculating the posterior probability, $P(c|x)$, from $P(c)$, $P(x)$, and $P(x|c)$. Naive Bayes classifier assume that the effect of the value of a predictor (x) on a given class (c) is independent of the values of other predictors. This assumption is called class conditional independence.

$$P(c \mid x) = \frac{P(x \mid c)P(c)}{P(x)} \tag{2.1}$$

$$P(c \mid x_j) = P(c_1 \mid x_j) * P(c_2 \mid x_j) * \cdots\cdots * P(c_n \mid x_j) \tag{2.2}$$

- P(c|x) is the posterior probability of class (target) given predictor (attribute).
- P(c) is the prior probability of class.
- P(x|c) is the likelihood which is the probability of predictor given class.
- P(x) is the prior probability of predictor.

2.5 Kappa Statistic

Kappa statistic is a generic term for several similar measures of agreement used with categorical data. Typically it is used in assessing the degree to which two or more raters, examining the same data, agree when it comes to assigning the data to categories[4]. It is frequently used to test inter-rater reliability. The importance of rater reliability lies in the fact that it represents the extent to which the data collected in the study are correct representations of the variables measured. Measurement of the extent to which data collectors (raters) assign the same score to the same variable is called inter-rater reliability. While there have been a variety of methods to measure inter-rater reliability, traditionally it was measured as percent agreement, calculated as the number of agreement scores divided by the total number of scores. In 1960, Jacob Cohen [4] critiqued use of percent agreement due to its inability to account for chance

[4] http://www.statistics.com/glossary&term_id=635

agreement. He introduced the Cohen's kappa, developed to account for the possibility that raters actually guess on at least some variables due to uncertainty. Like most correlation statistic, the kappa can range from −1 to +1. While the kappa is one of the most commonly used statistic to test inter-rater reliability.

2.5.1 Weighted Kappa Statistic

Weighted Kappa is a measure of agreement for categorical data. It is a generalization of the kappa statistic to situations in which the categories are not equal in some respect - that is, weighted by an objective or subjective function[5]. Weighted kappa lets you count disagreements differently and is especially useful when codes are ordered. Three matrices are involved, the matrix of observed scores, the matrix of expected scores based on chance agreement, and the weight matrix. Weight matrix cells located on the diagonal (upper-left to bottom-right) represent agreement and thus contain zeroes. Off-diagonal cells contain weights indicating the seriousness of that disagreement. Often, cells one off the diagonal are weighted 1, those two off 2, etc. The weighted kappa statistic provides a measure of agreement k_w*(scorer1, scorer2)* between two scorers *(scorer1 and scorer2)* who classify observations into one of several groups or categories. The following description is adapted from reference [1].

Two scorers (*scorer1 and scorer2*) analyze a set of N observations by classifying each observation into one of g groups. This leads to a $g \times g$-matrix n, the table of occurrences. A cell n_{ij} represents the number of observations that have been classified as belonging to category i by *scorer1* and to category j by *scorer2*. Depending on the particular situation to be investigated, a weight w_{ij} between zero and 1 is given to each cell n_{ij}. The weight w_{ij} quantifies the degree of discrepancy between the two categories i and j. Cells on the diagonal of the table of occurrences n, corresponding to equal classification by both scorers, are given weights of 1; whereas cells n_{ij} with highly different categories i and j receive weights w_{ij} close to or equal to zero. The weights w_{ij} given to the different cells n_{ij} of this second-by-second table of occurrences are displayed in Table 2.1.

[5] http://www.statistics.com/index.php?page=glossary&term_id=679

Table 2.1: Weights used for weighted kappa statistic to evaluate second-by-second agreement between two scorers of a PSG recording

	Scorer 2					
		N	H	OA	MA	CA
Scorer 1	N	1	0.5	0	0	0
	H	0.5	1	0.5	0.5	0.5
	OA	0	0.5	1	0.5	0.25
	MA	0	0.5	0.5	1	0.5
	CA	0	0.5	0.25	0.5	1

Whereas weights w_{ij} close to 1 indicate that the respective classes i and j are close, weights close to zero correspond to classes that are very different. On the one hand, the choice of this particular set of weights is motivated by the intention to penalize a missed apnea (weight zero) more severely than a missed hypopnea (weight 0.5) or than a respiratory event detected by both scorers but classified as an apnea by one of them and as a hypopnea by the other (weight 0.5). On the other hand, this set of weights specifically disfavors episodes classified as central apneas by one scorer and as obstructive apneas by the other (weight 0.25).

The weighted observed proportional agreement between the two scorers is obtained as:

$$P_{o(w)} = \frac{1}{N}\sum\nolimits_{i=1}^{g}\sum\nolimits_{j=1}^{g} w_{ij}n_{ij} \tag{2.3}$$

Abbreviating the row and column totals of the table of frequencies for the i^{th} category by $r_i = \sum_{j=1}^{g} n_{ij}$ and $c_i = \sum_{i=1}^{g} n_{ij}$, the weighted proportional agreement expected just by chance is estimated by

$$P_{o(w)} = \frac{1}{N}\sum\nolimits_{i=1}^{g}\sum\nolimits_{j=1}^{g} w_{ij}n_{ij}c_j \tag{2.4}$$

Then, weighted kappa, which may be interpreted as the chance-corrected weighted proportional agreement, is given by

$$k_w = \frac{P_{o(w)} P_{e(w)}}{1 - P_{e(w)}} \qquad (2.5)$$

It has a maximum of 1 when agreement between the two scorers is perfect; whereas a value of zero indicates no agreement better than chance, and negative values show worse than chance agreement.

Chapter 3
Measuring Opinion Credibility in Twitter

Our research is adding weight of tweets based on authors' background knowledge and calculating credibility of tweets polarity in twitter. Thereafter we converse all of our defined author's expertise weighted factors to analyze which weighted factors weightier than others and which factors can support more credible when we compute opinion credibility. We assume when we classify tweet content's polarity we think this is not fully credible result of tweet polarity. For example: Hong Kong Revolution tweets:

- **Dana Ryan**: More tourists are flocking to Hong Kong to watch the prodemocracy protesters, so business is good here. Merchants are happy.

- **RealHKNews**: #UmbrellaRevolution people's latest slogan "I want to go shopping". Picture showing disappointed shoppers sitting... http://t.co/VajNuZThnh

In this case, these two tweets intent to negatively affect Hong Kong business view on revolution issues. For example, whose opinion is more credential than others? To achieve that kinds of issue, we add weight of author's background knowledge for a given topic to measure credibility of tweets polarity on the following information:

- Who represents about this tweet? Or who is an author of these tweets?
- How much does this author know about this topic?
- Are they expert for this topic?
- What is their background knowledge?

In order to assess credibility of opinions, there are two main components in our system: sentiment analyzer and credibility calculator in Figure 3.1. Figure 3.1 shows our overview system to find credibility result of tweet polarity.

3.1 Preprocessing

In preprocessing step, we apply our method, we make preprocessing steps: remove non-English word (using WordNet to determine whether this word is English or not), remove url, # and @ (prefix, suffix), remove stop words, stemming, case folding in.

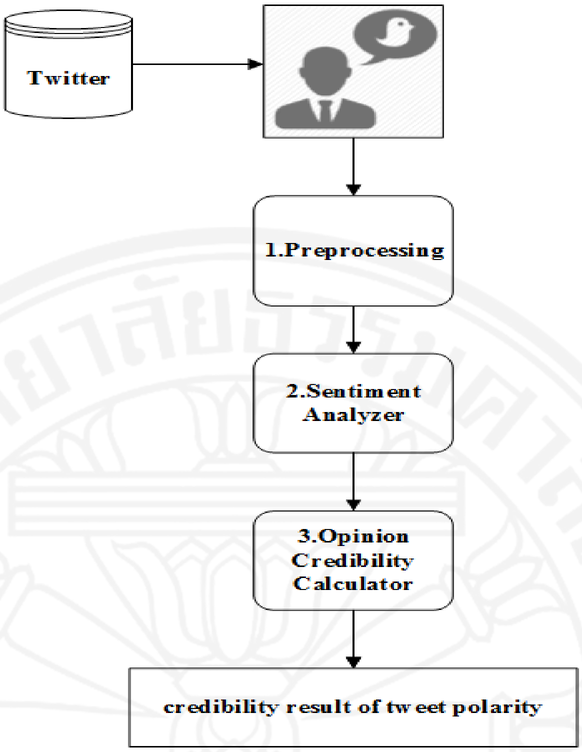

Fig 3.1. Overview system of credibility result of tweet polarity

3.2 Sentiment Analyzer

The task of sentiment analyzer is to achieve classification of tweets' polarity result. We perform sentiment classification to analyze which tweet is positive or negative using machine learning approaches: support vector machine (SVM) and Naïve Bayes (NB) to use which one is compatible for our specific topic.

Naïve Bayes is based on Bayesian theorem and one of the most basic text classification techniques. It analyze all the features in feature vector individually and these features are independent of each other. The conditional probability theorem is shown in equation (3.1). X is defined as feature vector and y_i is the class label.

$$P(X \mid Y) = \prod_{i=1}^{m} P(x_i \mid y_i) \quad (3.1)$$

SVM is a supervised learning method used for classification. SVM finds the hyperplane and separates two classes' points with the maximum margin [5]. We use

Linear SVM with Rapid Miner[6] software to classify sentiment polarity of our tweet dataset. w represents to find maximum hyperplane and separates document form one class or other. c_i corresponds positive and negative with class of document d_i. α_i is to solve dual optimization problem.

$$w = \sum_i \alpha_i c_i d_i, \alpha_i \geq 0 \tag{3.2}$$

3.3 Opinion Credibility Calculator

We compute opinion credibility based on an author's expert knowledge for a specific topic. In twitter, we consider twitter features: author's bio, List feature and author tweets behavior to identify weight of authors' expert knowledge.

- Author's bio: contains important information that indicates the expertise of the author, such as his/herself summarized interests, career information and links to his/her personal web page.
- Lists: allow users to organize people they are following into labeled groups. It contains self-reported expertise indicator, lists reflect external. i.e., follower's judgment about one's expertise and provide straightforward cues about this judgment to other users.
- Author's tweets: They tweeted almost everything: such as: their daily activities, comment on news, promotion of their company, etc.,

In List feature: name and description indicate valuable semantic cues about who has users included in the Lists are including their topics of expertise [3]. Based on Sharma N. K. et al. [3] approach, we use author's bio and List name, and description to infer author's background knowledge for a given topic. First we apply common language processing approaches: such as remove non-English word (using the WordNet to determine whether the word is English or not), remove url, # and @ (prefix, suffix), remove stop words, stemming, case folding. We apply N-gram approach (using unigram and bi-gram) to segment them and extract noun and adjective that are useful for charactering authors using NLP toolkit[7]. We used the ontology concept to filter a set of given topic to extract conceptual keyword that are related with given topic and

[6] www.rapidminer.org
[7] www.nltk.org

calculate the ratio of number of related keywords to total number of raw keywords. This strategy produces first score W_{BL} (author expert score using bio and List) to define the author's expert knowledge.

$$W_{BL} = \frac{\text{number of related keywords for a given topic}}{\textit{total}\ \text{number of raw keywords}} \tag{3.3}$$

Next step is to discover the expert score based on an author tweets behavior and his/her social network activities. We focus on the following features that reflect the impact of user background knowledge (expert knowledge score) for a given topic when we calculate the tweet's credibility;

- author's topic related ratio (TR)
- ratio of author's tweet retweet by other users (RT) for a given topic
- ratio of author's friends and followers who are related with given topic (F1, F2)
- author's opinions ratio for given topic (OP)

We combine these features to make the expert score for a given topic. For author topic related ratio, we make the ratio of number of author's tweet related for specific topic to the number of his/her all tweet (W_{TR}).

$$W_{TR} = \frac{\text{number of author's related tweets for a given topic}}{\text{number of author's all tweet}} \tag{3.4}$$

In Twitter, retweet is a re-posting someone else's Tweet. Using retweet features we compute how much times author's tweet has been retweeted by others (W_{RT}).

$$W_{RT} = \frac{\text{number of author's tweet retweet by others for a given topic}}{\textit{all}\ \text{author tweets for a given topic}} \tag{3.5}$$

W_{F1} and W_{F2} indicate the ratio of author's friends and followers who are related with given topic.

$$W_{F1} = \frac{\text{number of friends for a given topic}}{\text{total number of author's friends}} \tag{3.6}$$

$$W_{F2} = \frac{\text{number of followers for a given topic}}{\text{total number of author's followers}} \tag{3.7}$$

For author's opinions ratio; we assume how many times author expresses this opinions based on his/her past all opinions for given topic. e.g., given tweet opinion is negative, we calculate number of author's negative tweet based on his/her all past tweet opinion (*W_{OP}*).

$$W_{OP} = \frac{\text{number of author's negative or positive tweets}}{\textit{all} \text{ past opinions of author's tweets}} \tag{3.8}$$

To produce expert score, we combine all of these features by adding author's bio, List feature and his/her tweet behavior. Finally we compute expert score for given topic by adding author's bio, List features and author's tweet behavior.

$$\textit{Expert} \text{ score} = \frac{W_{BL} + W_{TR} + W_{RT} + W_{F1} + W_{F2} + W_{OP}}{6} \tag{3.9}$$

As the final result, we combine that expert score with the result of polarity from sentiment analyzer and calculate the credibility tweet polarity result *C (polarity)*.

$$C_{OP} = \text{tweet's polarity} \times \text{ Expert score} \tag{3.10}$$

3.4 Conversion of Weighted Factors

In this part, we compute which weighted factors are weightier than other factors when we identify credibility of opinion. Before changing weighted factors in expert score, we assume all of weighted factors have equal weight. Our purpose is conversion of each weighted factors and determine which factors is weightier than others. We compute Cop (Opinion Credibility) according to their value and evaluate with using weighted kappa statistic [15] based on the following algorithm. We identify which weighted factors is more influential than other factors comparing with credibility of opinions result.

Conversion of Weighted Factors Algorithm

1) Set W_{BL}=1 and (W_{TR}= W_{RT} = W_{F1}=W_{F2} = W_{OP}= 0); Find {C_{OP} =?}
2) Set W_{TR}=1 and (W_{BL}= W_{RT} = W_{F1}=W_{F2} = W_{OP}= 0); Find {C_{OP} =?}
3) Set W_{RT}=1 and (W_{BL}= W_{TR} = W_{F1}=W_{F2} = W_{OP}= 0); Find {C_{OP} =?}
4) Set W_{F1}=1 and (W_{BL}= W_{TR} = W_{RT}=W_{F2} = W_{OP}= 0); Find {C_{OP} =?}
5) Set W_{F2}=1 and (W_{BL}= W_{TR} = W_{RT}=W_{F1} = W_{OP}= 0); Find {C_{OP} =?}
6) Set W_{OP}=1 and (W_{BL}= W_{TR}= W_{RT}= W_{F1}= W_{F2}= 0); Find {C_{OP} =?}
7) Based on the result, classify the highest and lowest weighted factors.
8) Set two highest weighted factors' value to 0.5 and the rest to 0.
    a. Find {C_{OP} =?}
9) Decrease first two highest weighted factors' value from 0.5 to 0 by 0.1 and increase second two highest weighted factors' value from 0 to 0.5 by 0.1.
    a. Find {C_{OP} =?}
10) Decrease first two highest weighted factors' value from 0.5 to 0 by 0.1 and increase the lowest weighted factors' value 0 to 0.5 by 0.1.
    a. Find {C_{OP} =?}
11) Compare result of step 8, step 9 and step 10 and find which factors are weightier than others.

Chapter 4
Experimental Results and Discussion

4.1 Experiment Results with HongKong Dataset

We use Twitter API to collect data and crawl HongKong (HK) revolution data in the period between Sep 30, 2014 and Nov 30, 2014 as our training set and testing set. We use 1000 sample data for training dataset to learn sentiment classification using RapidMiner. We run this dataset with sentiment analyzer and show the output result, which labeled with tweet's polarity (positive, negative).

Table 4.1. Training Dataset

Dataset	positive	negative
Training Data	500	500
Test Data	100	100

In order to evaluate the accuracy of these polarity results, we use Linear SVM and calculate precision and recall. The following table shows the calculation results. We get precision 84.40% for positive and 96.93% for negative. We think rest of the errors is emoticons. According to the SVM classifier, it can occur more weight and reduce accuracy. We use many set of SVM parameters for finding the best result and we apply 10 fold cross validation and the best results are shown in Table 4.2.

Table 4.2. Result of Training Classification of HK Dataset

	true positive	true negative	class precision
positive	487	90	84.40%
negative	13	410	96.93%
class recall	97.40%	82.00%	

The second part of experiment is opinion credibility calculation. We take one sample tweet from training set. First, we compute this tweet polarity by using sentiment analyzer. Then we calculate weights and corresponding expert score for that tweet according to our equation 3.3 to 3.10. Finally, we get tweet polarity from sentiment analyzer and expert score, then we calculate credibility of tweet polarity. To evaluate credibility of tweet polarity, we calculate accuracy of opinion credibility. To do so we define author's expertise level as maximum, medium and minimum trusted levels. We manually categorizes the author of HK revolution dataset as that three level. Accordingly we set range for the credibility value as highest (h), middle (m) and lowest (l) range. This range setting depends on domain and classifier value. In our evaluation we assume that (h >=70%), (l <=30%), (h<=m >=l) respectively.

Then we determine the range of our classification result and verify with the author's expertise categories. From our proposed method, the opinion credibility of author [csn216] and [dw11138375] are 29% and 17%. Likewise, as we can see in Table 3, the author [csn126] is in the minimum trusted author group. In addition author [Dana_Ryan]: 36%, [lailaoshi]: 33% and [natashkakhanhk]: 68% are in medium trusted group and author [krisic]: 72% is maximum trusted group. From this comparison, we can conclude that our proposed method can provide reliable credibility.

Author's tweet: [csn126] Gotta love this guy. #occupyhk http://t.co/xg28uAjhYf

$$C(polarity) = tweets\ polarity \times expert\ score$$

tweet polarity : Positive: 1

w_{bl}= 0.55; w_{TR}=0.02; w_{RT}=0.0001; w_{F1}=0.25; w_{F2}=0.08; w_{OP}=0.81

Expert score = 0.55+0.02+0.0001+0.25+0.08+0.81/6 = 29%

C(polarity) = 1 (positive) x 29 = 29%

Fig.4.1. Example of how to measure opinion credibility

Table 4.3. Example of trusted Author group

Maximum	Medium	Minimum
krisic	lailaoshi	csn126
Ramyinocencio	Dana_Ryan	HKAYPGOLD
JeromeTaylor	natashkakhanhk	RealHKnews
fion_li	jen1113	dw11138375

Table 4.4. Sample Results of opinion credibility

No.	Authors	Tweets	Opinion	Opinion Credibility
1.	Dana_Ryan	It's so bad that 82% of Hong Kongers wanting to kill CY Leung for him playing the "Waiting Game". The police may not enforce the deadline.	negative	36%
2.	dw11138375	Heavy police presence and a lot of angry shoppers #OccupyCentral #UmbrellaRevolution #UMHK https://t.co/efPi74cvWF	negative	17%
3.	lailaosh	More tourists are flocking to Hong Kong to watch the pro-democracy protesters, so business is good here. Merchant is happy.	negative	33%
4.	krisic	"I am proud to be a HKer" @rosetangy #OccupyHK	positive	72%
5.	natashkakhan hk	"The most powerful weapon in winning democracy for HongKong is the people of the Umbrella Generation" -Benny Tai OpEd http://t.co/QN5bctyJ2y	positive	68%
6.	csn126	Gotta love this guy. #occupyhk http://t.co/xg28uAjhYf	positive	29%

4.2 Experiment Results with Obamacare Dataset

We crawl data from Obamacare (unofficial name) that is an American Government's Healthcare policy. To classify tweets' polarity, we use RapidMiner tool and 2000(training dataset) and 200(testing data). Our Sentiment Analyzer produce labelling tweet's polarity: positive and negative. We use Naïve Bayes classifier and Linear SVM classifier. Naïve Bayes is slightly lower accuracy than SVM and the accuracy results are 88.25% for Naïve Bayes and 90.50% for SVM.

Table 4.5: Results of training classification for Naïve Bayes (NB)

polarity	true positive	true negative	class precision
positive	925	160	85.25%
negative	75	840	91.80%
class recall	92.50%	84.00%	

Table 4.6: Results of training classification for SVM

polarity	true positive	true negative	class precision
positive	920	110	89.32%
negative	80	890	91.75%
class recall	92.00%	89.00%	

To evaluate the accuracy of opinion credibility, we used weighted kappa statistic [4] to assess how much rank for the agreement between our system and human raters. We have three ranges for our opinion credibility values such as: highest (h>=70%), lowest (l<=30%) and middle (h<=m>=l).

To determine the value of kappa statistic strength, we use Landis J. R. and Koch G. G. [2] approach. In our evaluation, R = {highest (h), lowest (l) and middle (m)}, N=100 for each positive and negative. For the matrix w follows as:

Table 4.7. Levels of agreement measure for Kappa statistic (K_w)

Kappa Statistic (Kw)	Strength of Agreement
<0.000	Poor
0.000-0.20	Slight
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Substantial
0.81-1.00	Almost Perfect

$$w = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix} \quad (4.1)$$

Table 4.8. Number of Agreement between system and rater for Positive Result

System	Rater			
	h	m	l	Total
h	18	7	3	28
m	6	9	11	26
l	8	6	32	46
Total	32	22	46	100

Table 4.9. Number of Agreement between system and rater for Negative Result

System	Rater			
	h	m	l	Total
h	8	10	6	24
m	7	12	13	32
l	4	6	34	44
Total	19	28	53	100

After we compute evaluation for no weight, we get result {positive: 0.60 (Moderate), negative: 0.52 (Moderate) and then evaluate also for conversion of our six weighted factor values for no weight in Table 4.10.

Step 1: Set one factor to '1' and the rest to '0'

Firstly, we set each of weighted factor stage to "0", expect one factor and give that factor stage to "1". E.g., IF (W_{BL} (stage) =1 (W_{TR}=W_{RT}=W_{F1}=W_{F2}=W_{OP}= 0) THEN C_{OP}=? In this step, we find W_{BL} and W_{TR} are the first highest weighted factors, W_{RT} and W_{F2} is the second highest weighted factors and W_{F1} and W_{OP} are the lowest weighted factors among these six weighted factors in Table 4.10: Step 1.

Table 4.10. Result of kappa statistic (K_w) between System and Rater for each conversion of weighted factors

	Weighted Factors						Result of Kappa Statistic (Kw)	
	W_{BL}	W_{TR}	W_{RT}	W_{F1}	W_{F2}	W_{OP}	Positive	Negative
No Weight	1	1	1	1	1	1	0.60 (Moderate)	0.52 (Moderate)
Step 1	0	1	0	0	0	0	0.53(Moderate)	0.49(Moderate)
	1	0	0	0	0	0	0.42(Moderate)	0.44(Moderate)
	0	0	1	0	0	0	0.37(Fair)	0.21(Fair)
	0	0	0	0	1	0	0.34(Fair)	0.24(Fair)
	0	0	0	1	0	0	0.18(Slight)	0.17(Slight)
	0	0	0	0	0	1	0.07(Slight)	0.08(Slight)
Step 2	0.5	0.5	0	0	0	0	0.58(Moderate)	0.50(Moderate)
Step 3	0.4	0.4	0.1	0	0.1	0	0.47(Moderate)	0.46(Moderate)
	0.3	0.3	0.2	0	0.2	0	0.38(Fair)	0.34(Fair)
	0.2	0.2	0.3	0	0.3	0	0.19(Slight)	0.17(Slight)
	0.1	0.1	0.4	0	0.4	0	0.09(Slight)	0.11(Slight)
	0	0	0.5	0	0.5	0	0.32(Fair)	0.33(Fair)
Step 4	0.4	0.4	0	0.1	0	0.1	0.17(Slight)	0.15(Slight)
	0.3	0.3	0	0.2	0	0.2	0.15(Slight)	0.12(Slight)
	0.2	0.2	0	0.3	0	0.3	0.09(Slight)	0.08(Slight)
	0.1	0.1	0	0.4	0	0.4	0.08(Slight)	0.07(Slight)
	0	0	0	0.5	0	0.5	0.05(Slight)	0.07(Slight)

Step 2: Set two highest weighted factors' value to 0.5 and the rest to 0

Based on the step 1 experiment result, we notice W_{BL} and W_{TR} are the first highest weighted factors in our weighted factors. Therefore we give these two highest weighted factors to {W_{BL} =0.5, W_{TR} =0.5} and set 0.0 for the rest and calculate accuracy of our opinion credibility result again in Table 4.10: Step 2.

Step 3: Decrease first two highest weighted factors' value (W_{BL}, W_{TR}) and increase second two highest weighted factors' value (W_{RT}, W_{F2})

In this step, we decrease two highest weighted factors' values (W_{BL} and W_{TR}) from 0.5 to 0 by 0.1. Simultaneously, we also increase the second two highest weighted factors' values (W_{RT}, W_{F2}) from 0 to 0.5 by 0.1. The results are in the following Table 4.10: Step 3.

Step 4: Decrease first two highest weighted factors' value (W_{BL}, W_{TR}) and increase two lowest weighted factors' value (W_{F1}, W_{OP})

In step 4, we decrease two highest weighted factors' values (W_{BL} and W_{TR}) from 0.5 to 0 by 0.1. Simultaneously, we also increase two lowest weighted factors' values (W_{F1}, W_{OP}) from 0 to 0.5 by 0.1. The results are in the following Table 4.10: Step 4.

Based on our conversion of weighted factors value, we find that W_{BL} and W_{TR} are the most important weighted factors in our six factors because we discover our accuracy result values are decreased when we reduce their values (W_{BL} and W_{TR}).

Chapter 5
Conclusions

Sentiment Analysis is a type of natural language processing for tracking the attitudes, feelings or appraisal of the public about particular topic, product or services. In previous researches, they used many methods or ways to analyze human opinions. At that point, we noticed to identify credibility of opinion results. How do we know credibility of opinions?

In this thesis, we proposed to calculate credibility of opinion using expert knowledge in one of the social media (Twitter). In twitter, we use author's profile, twitter lists feature and their tweets' behaviour for expert knowledge. Based on these features, we specify six weighted factors: W_{BL} (using user bio and lists features), W_{TR} (author's topic related ratio), W_{RT} (ratio of author's tweet retweet by other users (RT)), W_{F1} (ration of author's friends), W_{F2} (ratio of author's followers) and W_{OP} (authors' opinions ratio for given topic. Next we classify a given tweet as either positive or negative and we called this part is Sentiment Analyzer in our research and we combine our defined six factors and we called this is expert score. Thereafter we compute credibility of tweet polarity result *Cop* (Opinion Credibility) using tweet's polarity result from Sentiment Analyzer and Expert score. Moreover we converser all of our weight value and compute which weighted factors are more influential than other factors.

By experiments, we used two dataset HongKong revolution (HKrevolution) data and Obamacare data for our research. For HKrevolution data we used 1000 sample data for training dataset and 100 data for testing data for sentiment classification. We get precision (84.4%) for positive and 96.93% for negative using Support Vector Machine (SVM) classifier. For Obamacare data we used 2000 (training) and 200 (testing) data and got accuracy for 88.25% for Naïve Bayes classifier) and 90.50% for SVM classifier respectively.

To evaluate accuracy of opinion credibility, we used weighted kappa statistic to estimate how much rank for the agreement between our system and human raters. We get result {positive; 0.60 (Moderate), negative: 0.52 (Moderate)}. And then evaluate also conversion of our six weighted factor values and we find some factors are important and give more value for opinion credibility result. Based on our result, W_{BL} (author description and list feature) and W_{TR} (author topic ratio) are the most important weighted factors in our six factors for a given domain because we discover our result values are decreased when we reduced their values and they influence other factors values. We notice using weighted kappa statistic for evaluation is not good for real-time applications. For future work, we will focus to modify our expert score calculation based on conversion of our weighted factors result with different evaluation technique in other topic.

References

1. Altman, D. G. (1991). *Practical Statistic for Medical Research*. Boca Raton, FL: Chapman & Hall.

2. Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometric*, *33*, 159-174.

3. Sharma, N. K., Ghosh, S., Benevenuto, F., Ganguly, N., & Gummadi, K. (2012). *Inferring who-is-who in the Twitter social network*. Paper presented at the Proceedings of the 2012 ACM workshop on Workshop on online social networks, Helsinki, Finland.

4. Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological bulletin*, *70*(4), 213-220.

5. Hsu, C., Chang, C., & Lin, C. (2010). *A practical guide to support vector classification*. Retrieved from December 19, 2015, http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf

6. Hu, M., & Liu, B. (2004). *Mining and summarizing customer reviews*. Paper presented at the Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, Seattle, WA, USA.

7. Garcia-Moya, L., Anaya-Sanchez, H., & Berlanga-Llavori, R. (2013). Retrieving Product Features and Opinions from Customer Reviews. *IEEE Intelligent Systems*, *28*(3), 19-27.

8. Bafna, K., & Toshniwal, D. (2013). Feature based Summarization of Customers' Reviews of Online Products. *Procedia Computer Science*, *22*, 142-151.

9. Pang, B., Lee, L., & Vaithyanathan, S. (2002). *Thumbs up?: sentiment classification using machine learning techniques*. Paper presented at the Proceedings of the ACL-02 conference on Empirical methods in natural language processing - Volume 10.

10. Pang, B., Lee, L., & Vaithyanathan, S. (2010). *Twitter as a Corpus for Sentiment Analysis and Opinion Mining*. Paper presented at the Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10).

11. Neethu, M. S., & Rajasree, R. (2013, July). *Sentiment analysis in twitter using machine learning techniques*. Paper presented the Processding of the Fourth International Conference on Computing, Communications and Networking Technologies.

12. Liang, P.-W., & Dai, B.-R. (2013). Opinion Mining on Social Media Data. 91-96.

13. Go, A., Bhayani, R. & Huang, L. (2009). *Twitter Sentiment Classification using Distant Supervision*. Retrieved from December 19, 2015, https://cs.stanford.edu/people/alecmgo/papers/TwitterDistantSupervision09.pdf

14. Liu B. (2012). *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers.

15. Booker, M. J., & McNamara, A L., (2004). Solving black box computation problems using expert knowledge theory and methods. *Reliability Engineering & System Safety*, *85*(1-3), 331–340.

16. Kaplan, S. (1992). 'Expert information' versus 'expert opinions'. Another approach to the problem of eliciting combining using expert knowledge in PRA. *Reliability Engineering & System Safety*, *35*(1), 61-72.

17. Ayyub, B. (2001). *Elicitation of expert opinions for uncertainty and risks*. Boca Raton, FL: CRC Press.

18. Weng, J., Lim, E.-P., Jiang, J., & He, Q. (2010). *TwitterRank: finding topic-sensitive influential twitterers.* Paper presented at the Proceedings of the third ACM international conference on Web search and data mining, New York, New York, USA.

19. Pal, A., & Counts, S. (2011). *Identifying topical authorities in microblogs*. Paper presented at the Proceedings of the fourth ACM international conference on Web search and data mining, Hong Kong, China.

20. Liang, C., Liu, Z., & Sun, M. (2012, December). *Expert Finding for Microblog Misinformation Identification*. Paper presented at the Proceedings of COLING 2012, Mumbai, India.

21. Namihira, Y., Segawa, N., Ikegami, Y., Kawai, K., Kawabe, T., & Tsuruta, S. (2013). High Precision Credibility Analysis of Information on Twitter. 909-915.

Appendix
List of Publications

A.1 International Conference

1. Mya Thandar, Sasiporn Usanavasin, “Measuring Opinion Credibility in Twitter”, Proceedings In 11th International Conference on Computing and Information Technology (IC2IT), Recent Advances in Information and Communication Technology 2015 ,Advances in Intelligent Systems and Computing Volume 361, 2015, pp 205-214

2. Mya Thandar, Sasiporn Usanavasin, “Weighted Factors for Opinion Credibility in Twitter”, Proceedings in Innovative Research in Engineering and Technology (IRET-16), Jan 21-22, 2016.