Journal of Universal Computer Science, vol. 18, no. 8 (2012), 973-992
submitted: 22/9/11, accepted: 14/12/11, appeared: 28/4/12 © J.UCS

Discovering Consumer Insight from Twitter via Sentiment Analysis

Wilas Chamlertwat, Pattarasinee Bhattarakosol, Tippakorn Rungkasiri
(Chulalongkorn University, Bangkok, Thailand
wilas@chamlertwat.in.th, pattarasinee.b@chula.ac.th, tippakorn@acc.chula.ac.th)

Choochart Haruechaiyasak
(National Electronics and Computer Technology Center, Pathumthani, Thailand
choochart.haruechaiyasak@nectec.or.th)

Abstract: Traditional approaches for studying consumer behavior, such as marketing survey and focus group, require a large amount of time and resources. Moreover, some products, such as smartphones, have a short product life cycle. As an alternative solution, we propose a system, the *Micro-blog Sentiment Analysis System* (MSAS), based on sentiment analysis to automatically analyze customer opinions from the Twitter micro-blog service. The MSAS consists of five main functions to (1) collect Twitter posts, (2) filter for opinionated posts, (3) detect polarity in each post, (4) categorize product features and (5) summarize and visualize the overall results. We used the product domain of smartphone as our case study. The experiments on 100,000 collected posts related to smartphones showed that the system could help indicating the customers' sentiments towards the product features, such as *Application*, *Screen*, and *Camera*. Further evaluation by experts in smartphone industry confirmed that the system yielded some valid results.

Keywords: Micro-blog, sentiment analysis, natural language processing, information extraction, information visualization
Categories: H.3.5, H.4.3, M.0

1 Introduction

For the past few years, we have witnessed a flourishing of social media. People have been building a global communication network on the Internet via numerous social network applications. Everyday a large amount of unprecedented content is generated on social networks, especially on micro-blog services [Jung, 08]. The increasingly popular use of micro-blog for lightweight communication raises its potential for serving as a new communication medium [Jung, 09]. Twitter, one of the most popular micro-blog services, claimed that there are over 200 million Tweets or posts per day [Twitter blog, 11]. It is a revolution of how content is generated and distributed by creating, sharing, and discovering messages without controlling. Mark Zuckerberg, the Facebook founder, said that this behavior has become a new social norm [Cashmore, 10].

As consumers, most Internet users have already integrated social media into their daily lives. It is much easier and more accessible to share any information on any user devices and interfaces [Jung, 10a]. Moreover, the way we communicate has been changed. Consumers do not unilaterally believe messages from producers, but people need dialogue to convey their messages to the others. Obviously, Twitter has dramatically lowered the barriers to create content, and thus sharing day-to-day lives became effortless. People can update their status and share their opinions with friends and followers in the network any time anywhere [Jung, 10b].

As producers, private enterprises also widely adopt social media for their marketing strategy to commercialize their products or services, not only to speak, but also to listen to the true voice of the customers in their own words. Although most people are hesitated to answer survey about product or service preference, they express their thoughts in social network and wield enormous influence in shaping the opinions of other consumers [Zabin, 08]. Micro-blog is a rich resource for hearing this useful information. It contains many user sentiment expressions either positive or negative to many different topics in the market [Jung, 11a].

Compared to traditional website or blog, micro-blog is a social networking application that grows tremendous user opinions. Unlike a short-term technology trend, micro-blog is the evolution of mini blogging that many are latching on. Since communication in micro-blog is short and casual, people can keep updating on one another's activities in real time [Jung, 11b]. It fulfils today's user behavior for a fast and immediate manner. As a result, products or services appreciations can be easily found in micro-blogs. These messages can lead to the decision making for buying or ignoring of the consumers. The magnitude and swiftness of the influence of micro-blog challenges researchers and practitioners with a need to understand the science behind it and devise applications for marketing communications.

The analysis of large-scale social data, which confirms corporate perspective from consumer opinions, is really essential for supporting top-level management to solve the real-world problem. These consumer voices can influence brand perception, brand loyalty and brand advocacy. With social media monitoring and sentiment analysis, the enterprises will be able to tap into consumer insights to improve their quality of product, provide better service, or even identify new business opportunity, and other activities accordingly.

In smartphone market, the advancement of technology is continually surpassing one another and updated versions are hitting the market at lightening speeds. According to HTC, the average shelf life for smartphones has been decreasing from three years in 2007 to around six to nine months in 2011 [Ferreira, 11]. The producers have too little time to research market by traditional way. Referring to Technology Adoption Life Cycle, each model of smartphones has limited time to prove their product concept. We need an alternative way to retrieve necessary information to improve next generation product [Jung, 12].

Since there are many people exchange their information about smartphones on Twitter, sentiment analysis may be appropriate technique to discover consumer insight. For example Twitter post #1 says, "Oh I love the 'Voice Search' Feature in my new HTC Sensation. Saves me the trouble of typing long sentences." Twitter post #2 says, "I hate htc peep!!! UberSocial is hands downe da best cuz its faster". This example can hint HTC to keep some features while to terminate or improve some features.

This study will try to investigate whether the result from sentiment analysis on micro-blog is valid for understanding consumer insight in some industries, such as smartphone market, or not. We propose the *Micro-blog Sentiment Analysis System* (MSAS) to accomplish the task. Both machine learning and lexicon-based technique will be applied to acquire the appropriate result. The MSAS helps revealing the customers' opinions towards smart phones' features as positive or negative. The results are summarized and displayed in some visualization formats. The final results from the MSAS are evaluated by three experts in the smartphone industry.

The rest of this paper is organized as follows. In next section, we review some related research works in consumer insight and sentiment analysis. In Section 3, we describe our proposed system the MSAS with full detail. Section 4 presents the analyzed results from the system using a large set of Twitter posts. Section 5 concludes the paper with some notes on future work.

2 Related Works

2.1 Consumer Insight

Consumer insight is the study about who the consumer is and what they think or feel [Stone, 04]. It may be the revolution from direct marketing to database marketing to customer relationship management or customer experience management. Enterprise can utilize this understanding to support from changing their marketing strategies to improving their operation and interactions. Good consumer insight is the foundation of good customer relationship management. Therefore, enterprises will be able to position themselves as they want it to be to meet customers or stakeholders need.

Insight is not conscious behaviors or thoughts of consumers but most are affected by various external factors from the state of economy and society to the way a brand is marketed. Enterprise can collect this information from complaints or compliments or through requests for further information. Some tasks can achieve from market research while some need mining from customer database, feedback from sales and customer service staffs.

Although measuring customer satisfaction has become very significant, setting staff targets using information based on what consumers are saying can stifle enterprise's creativity and limit innovation. In large corporate, the measurement will mislead by reducing cost and increasing control. This focus has become internal; leading staff and the entire organization away from first considering the customer. As a consequence, they fall in customer loyalty and overall satisfaction. The consumer insight process helps to change how we think about consumers, staff and our organizations, rather than supports a counterproductive way of thinking about them.

Applications for mining large volumes of textual data for marketing intelligence can be categorized into three types [Glance, 05] as follows:

- Early alerting - informing subscribers when a rare but critical, or even fatal, condition occurs.
- Buzz tracking - following trends in topics of discussion and understanding what new topics are forming.
- Sentiment mining - extracting aggregate measures of positive vs. negative opinion.

Opinion Observer is an early framework for analyzing and comparing consumer opinions of competing products in the market [Liu, 05]. The system provides a single glance of its visualization. The users are able to clearly see the strengths and weaknesses of each product in the minds of consumers in terms of various product features. For a product manufacturer, the comparison enables it to easily gather marketing intelligence and product benchmarking information. Second, a new technique based on the language pattern mining is proposed to extract product features from Pros and Cons in a particular type of reviews.

2.2 Sentiment Analysis

Sentiment analysis, or so called opinion mining, has been studied by many researchers in recent years. Sentiment analysis is a type of computational study of text in natural language which aims to identify sentiment polarity, intensity, and topics those sentiments apply to [Liu, 11]. Sentiment analysis turns out of the need for automated opinion disclosure and summarization system dealing with the large amount of data to allow machine to understand human generated content [Lake, 11]. In business arena, this technology has been used in mining opinions for brand monitoring, polls, financial trading, marketing and many more real world problems.

Traditionally, enterprises may conduct a consumer survey to investigate what customers want. A number of biases arise during the survey. The bias problem is known as the Bradley effect [Bobo, 09]. People are unwilling to provide accurate answer that reflects unpopular attitudes or opinions. Even though the researchers come up with well-designed surveys that can provide quality estimations, they could be costly and time consumed. Sentiment analysis is an alternative to analyze existing data without bias by not explicitly asking any questions to the people. This approach will reflect people true opinion than traditional survey responds. However, there is a drawback regarding sampling population which may not certify the target group of responders. As a result, sentiment analysis would not be able to completely replace the traditional approach but it could work as a complimentary solution.

Technically, sentiment analysis is a part of Natural Language Processing (NLP) study. Subjectivity classification and sentiment classification are perhaps the most widely studied topics in this field. Subjectivity classification is a process to separate subjective from objective sentences or distinguish opinions from facts while Sentiment classification is a process to determine sentiment orientation whether that sentence expressed positive or negative feeling. Also, some researches are interested in determining the intensity (strength) of sentiment polarity to measure the semantic intensity. Feature-based sentiment analysis is an in-depth study that refers to the determining of the expressed sentiments on different features of entities. For example, the feature-based sentiment analysis of smart phone screen is the study about people expression on screen whether it is positive or negative.

Various sentiment analysis methods have been researched in several different levels of text granularity [Agarwal, 11]. The coarse-grain level starts from a document level classification task [Turney, 02] to a finer-grain level of a sentence [Hu, 04] and at the phrase level [Wilson, 05]. Fundamentally, the approaches used for sentiment analysis can be divided into two categories, machine learning approach and lexicon-based approach.

1) Machine learning approaches are supervised learning approaches. It is referred to as training process that teaches an agent to classify input to output. Once enough

training data labelled with sentiment values is learned by the algorithm, sentiment analysis on corresponding domain data will provide promising results [Pang, 02].

We briefly describe three popular classification algorithms:

- Naive Bayes (NB) is a simple probabilistic classifier based on applying Bayesian theorem and is particularly suited when the dimensionality of the inputs is high.
- Max Entropy (MaxEnt) or a multinomial logit model is commonly used as alternative to Naive Bayes. Particularly, learning in a Naive Bayes classifier is a simple matter of counting up the number of cooccurrences of features and classes, while in a maximum entropy classifier the weights, which are typically maximized using maximum a posteriori (MAP) estimation, must be learned using an iterative procedure.
- Support Vector Machines (SVM) has been highlighted as one of the best performance. It is the method that analyzes data and recognizes patterns, used for classification and regression analysis. The essence of SVM is to find a hyperplane that separates document vectors from one class to the other as much as possible.

2) Lexicon-based approaches are typically unsupervised approaches. It is a rule based on features provided by predetermined sentimental lexicon score to estimate the polarity whether it is positive or negative. These approaches can function without any reference corpus and preceding training. Sentiment lexicons usually correctly estimate the generic polarity of term in a way that does not take domain information into account [Zhe, 10].

Opinion words are words that are used to state positive or negative sentiment. Words that express desirable feeling such as great or excellence have a positive polarity while words that express undesirable feeling such as bad or awful have a negative polarity. The dictionary of opinion words to identify sentiment and its orientation is called “opinion lexicon”. Some of interesting opinion lexicon is *SentiWordNet* and *Pageranking WordNet*.

SentiWordNet is one of public linguistic resources that assign to each synset, or synonym set, of WordNet with three numerical scores: positivity, negativity, objectivity [Esuli, 07b]. Pageranking WordNet is an applicability of a random-walk model to the determination of WordNet synsets in terms of how strongly they possess a given semantic property [Esuli, 07a].

Sentiment analysis on micro-blog has received scholar intention recently. Micro-blog contains important information that state positive or negative feeling in very limited space, for example, “The htc hd2 is awesome, the 1GHz processor really helps.”

However, it is not always as easy as identifying “I love android” or “He hates iPhone.”. The language people speak today is complicated and full of slang, ambiguity, sarcasm, idiom, and irony as in the following example, “HTC battery die in 15 minutes but take a year to charge. I'm so lucky!”.

Early research on this issue may be twitter sentiment classification using distant supervision [Go, 09]. This work showed that machine learning algorithms have accuracy above 80% when trained with emoticon data and SVM outperforms other classification while unigram model surpass bigram and parts-of-speech (POS) features. Moreover, there is a study confirm that POS features are not useful for sentiment analysis in the micro-blog domain [Kouloumpis, 11].

Later interesting work is the combination of both lexicon-based and machine learning-based approaches. In this study, we use a lexicon-based approach to perform entity-level sentiment analysis, which can give high precision but low recall. Then, to improve recall and the F-score, machine learning-based approach is also applied to identify opinion automatically by exploiting the information in the result of lexicon-based method [Zhang, 11].

3 Proposed solution

In this paper, we propose an approach based on sentiment analysis to determine whether a micro-blog post is a positive or negative sentiment. We use Twitter to represent the micro-blog service. Both machine learning based and lexicon-based approaches are applied in our solution. In this section, we start by giving an overview of Twitter characteristics. Then, the proposed methodology will be given with a full detail.

3.1 Twitter characteristics

Twitter, owned and operated by Twitter Inc., is the most popular micro-blogging service among other existing equivalents, such as *Friendfeed*, *Tumblr*, and *Identi.ca*. Twitter users can post short messages, called *tweets,* on their user profile and read others' messages on a single list aggregated in a reverse chronologically ordered, called *timeline*. Tweets are text-based posts limited to 140 UTF-8 characters about any updates from small little things happening in user daily life. The short nature of updates allows users to post quickly in real-time, reaching their audience immediately. By default, tweets are publicly visible but the owner can set privacy to display only to their friends. The relationships between users, or so-called *following*, are asymmetric. User can follow others and see their tweets, but the other users need not reciprocate. The subscribers are known as *followers*. Two subscribers are friends when both of them mutually follow each other.

Users can interact with Twitter directly through the website *twitter.com*, and many third party applications, ranging from web-based applications, desktop clients, and mobile phones. Because Twitter provides Application Programming Interface (API) to allow the integration with other services, users can deliver their messages directly to followers via Short Message Service (SMS), Instant Messaging (IM), Really Simple Syndication (RSS), e-mail, and many other tools, including posts on behalf of the user from automated agents. The ecosystem around Twitter is massively extensive though worldwide developers.

According to the constraint of limited characters in each message, Twitter embraces a series of conventions to allow users to add structure to their tweets.

1) address

To reply or mention, users can refer other users in their tweets using the @ symbol in front of username (also known as @reply). For example, “@alice that's so awesome, congrats!” This action makes a link between users and allows for threaded conversations among users. Moreover, users can use the letter *d* followed by username to send a private message to a specific user.

2) hashtag

To freely categorize tweets together, users can label using of hashtags or # symbol in front of words. For example, “Can't wait the new #iphone4”. The practice of using hashtags may stem from HTML anchor point or prefacing specialized words with punctuation marks in computer programming [Boyd, 10].

3) external link

To share link in constricted space, users may use URL shortening service to generate a unique abbreviated URL that redirects to desired website. For example, http://bit.ly/17zKu comes from http://en.wikipedia.org/wiki/Microblogging. From March 2010, Twitter has provided t.co, link shortening service, for links posted in Twitter to protect users from malicious sites, and to track clicks on links within tweets.

4) retweet

To repost another users' message on Twitter, users can do this in two manners. First, traditional retweet (RT) is applied by copying the post and preceding it with RT @username. For example, Alice posts “Happy new year” then Bob retweets “RT @Alice: Happy new year”. Second, new Twitter retweet is much easier by clicking the retweet link in the web and the tweet will, then, be forwarded to all of user's followers.

Several studies have tried to analyze the usage of Twitter. Main types of user intentions are daily chatter, conversations, sharing information and reporting news while main categories of users are information source, friends and information seeker [Java, 07]. Micro-blog is one of generating conversation mechanism that complements other form of interaction [Zhao, 09]. Customers can use micro-blog as a tool for words of mouth communications and discussion while corporates can use micro-blog as part of their overall marketing strategy [Jansen, 09].

3.2 Methodology

One form of sentiment analysis in product reviews is to produce a feature-based summarization [Hu, 04]. Features of a product are attributes, components, and other aspects of the product such as size, power, and display. In this paper, we propose a solution, called the *Micro-blog Sentiment Analysis System* (MSAS), to analyze the sentiment on Twitter. Our proposed system, the MSAS, could determine positive or negative sentiments on product features; then, aggregate the results to produce a summary for the users.

As illustrated in Figure 1, the MSAS processes can be separated into two phases: the *Preparation* phase and the *Analysis* phase.

The Preparation phase is the stage that all required data, including posts, model, relevant lexicon, are arranged on a specific domain as follows.

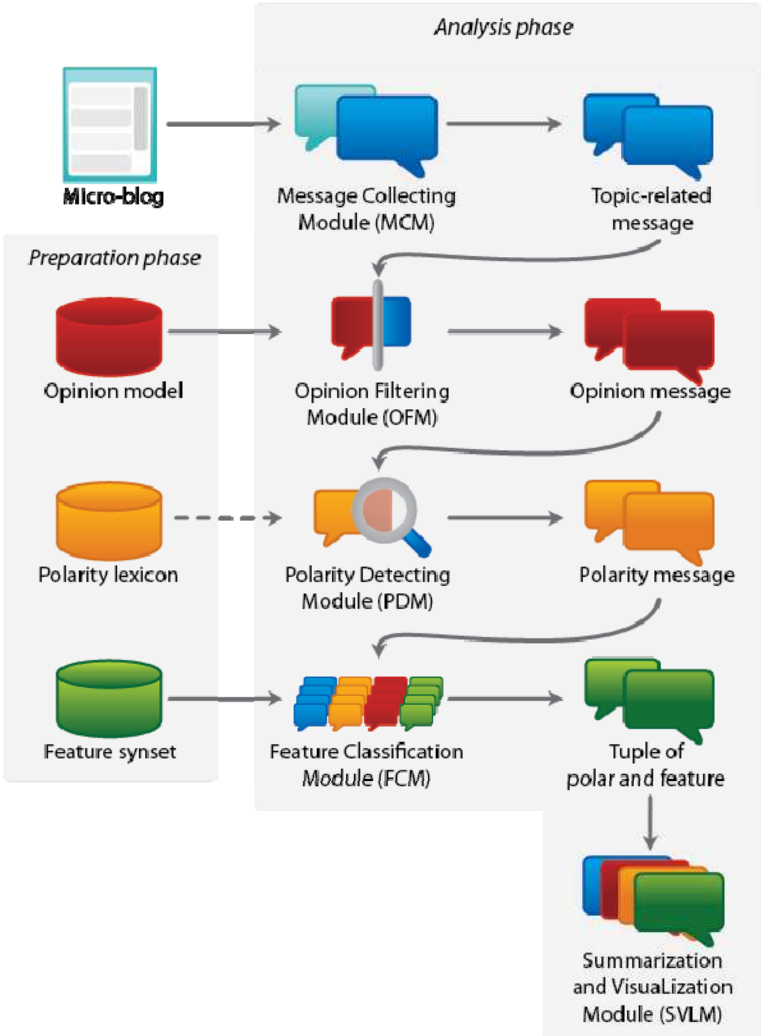

Figure 1: Micro-blog Sentiment Analysis System (MSAS)

1) Micro-blog: this is the input source to the system. As mentioned earlier, Twitter can be a good representative of micro-blog since it provides tremendous amount of data that is available as a public timeline for easy gathering. As illustrated in Table 1, we use the domain of *smartphones* as our case study.

id	created_at	message_status_id	text
1	2010-08-28 00:49:42	22285927850	I love this Samsung Galaxy but if u get one dont plug it in straight away to fully charge the battery. Instead let the battery run out...
2	2010-08-27 22:05:01	22274155397	Motorola Milestone has the 'best mobile display in the world'. Ref: displaymate.com
3	2010-08-28 02:32:26	22292491844	None of the three cameras I can videotape with focus on close up items. But my iPhone 4 does. How lame is that?

Table 1: Examples of collected tweets stored in database

2) Opinion model: this model is created based on classification algorithm. The model is used for filtering opinionated posts from non-opinionated ones. Table 2 shows some examples of tweets classified as *opinion* and *non-opinion* posts. The sentiment of each tweet will be further analyzed by using polarity lexicon.

Type	Sentiment	Message
opinion	Positive	wow that's awesome! Enjoy customizing. :) Nexus one is probably the father of customizations with several types of roms available
	Negative	My Nexus One is not responding well to location change for mobile network signal change. Need an app to manually refresh network.
non-opinion	N/A	Look for Verizon's new release of the Motorola Devour as well as an HTC Hero coming soon. AT&T to finally introduce Android OS!!!

Table 2: Examples of opinion and non-opinion tweets

3) Polarity lexicon: it is a dictionary used for the lexicon-based sentiment analysis approach. We obtain our opinion lexicon from SentiWordNet 3.0 [Baccianella, 10]. However, we discard objective terms since it does not contain positive or negative polarity. Table 3 shows some examples of terms from SentiWordNet 3.0. In this table, POS refers to Part of Speech; the pair (POS, ID) uniquely identifies a WordNet (3.0) synset. The values PosScore and NegScore are the positivity and negativity scores. SynsetTerms are set of synonymous terms with sense number written after the # symbol. Lastly, Gloss is meaning of that word.

POS	ID	PosScore	NegScore	SynsetTerms	Gloss
a	01586866	0.75	0	pleasant#2	(of persons) having pleasing manners or behavior; "I didn't enjoy it and probably wasn't a pleasant person to be around"
a	01587077	0	0.875	nasty#1 awful#3	offensive or even (of persons) malicious; "in a nasty mood"; "a nasty accident"; "a nasty shock"; "a nasty smell"; "a nasty trick to pull"; "Will he say nasty things at my funeral?"- Ezra Pound

Table 3: Data format of SentiWordNet v3.0

4) Feature synset: this is a collection of multiple product or service features that companies would like to study. Since different customers may refer to each feature differently, we have to group them into feature categories by studying the product specification across all manufacturers' websites. Table 4 shows some examples of feature categories and feature terms for smartphones.

Category	Features
Network	2g, 3g, connection, wifi, wireless, cellular
Screen	display, ppi, pixel, capacitive, resolution, monitor, lcd
Capacity	capacity, disk, storage, ram, rom, memory

Table 4: Examples of smartphone feature categories

Once we have already prepared everything, the Analysis phase is the examination stage composed of four related modules as follows.

1) Message Collecting Module (MCM):
This module utilizes Twitter Search API to collect tweets that we would like to study. Twitter has an Application Programming Interface (API) for programmatically accessing tweets by query term [Twitter Website, 11]. The search result is made anonymously without authentication. At the time we collect our data, Twitter API has a limit of 100 tweets in a response for any request. Data can be retrieved as XML or JSON format. For smartphones, we use query terms by following related keywords, such as iphone, nexus, htc, motorola, nokia, blackberry, and palm.

2) Opinion Filtering Model (OFM):
This module classifies a given tweet as "*subjective* or *opinion*" and "*objective* or *non-opinion*" using a machine learning approach. First, we started by collecting training data and labelling tweets manually as opinion or non-opinion. Then, we performed the following taks, feature selection and classification model construction. In our previous study, we found that the best performance for filtering opinion tweets is the using of the Support Vector Machines (SVM) with the information gain (IG) feature selection. The model yielded the accuracy equal to 84.5% [Chamlertwat, 11].

3) Polarity Detecting Module (PDM):
This module determines the polarity of a given message as positive or negative. It could also indicate the strength of the sentiment words. The module calculates polarity score for each tweet by averaging the scores of each sentiment words found in the tweet. We pre-process each Twitter post as follows:
a) tokenization – we segment each post by observing the word delimiters, such as space and punctuation marks. For example, a post "Just realised that the screen on the iPhone is the same size as my android phone" is tokenized to 16 words, "Just", "realised", "that", "the", "screen", "on", "the", "iPhone", "is", "the", "same", "size", "as", "my", "android", "phone". All tokenized words form a bag of words.
b) stopword removal – we remove stopwords from the bag of words by looking up a stopword dictionary. Stopword removal helps reduce term dimension by cleaning up unmeaningful terms.
c) link removal – we remove address and external link, since they are not useful in detecting polarity and features.
d) term normalization – we change the form of verb and adjective back to base form, and replace the abbreviation with its meaning. For example, the verb "talking" is replaced with "talk".
e) slang handling – we replace a sequence of repeated characters by detecting a sequence of three or more characters, for example, the word “gooooooood” is reduced to “good”
Then, we score each tweet by applying these rules:

(1) initialize the total polarity score: s ← 0
(2) check each token with SentiWordNet

- if token is positive, then s ← s + w; where w is score of sentiment word
- if token is negative, then s ← s – w

(3) if tweet contains negation, then s ← -s; where negation are the word “no” or “not” that will twist the overall tweet polarity

4) Feature Classification Module (FCM):
This module is responsible for outlining the brand or the product feature of each tweet. For example, “Twitterific is pretty great. I do all my tweeting through my iPhone”. This tweet is classified as follows:

twitterific, tweeting → feature
pretty → sentiment (score = 0.20833333333333334)
great → sentiment (score = 0.10693039146587664)
iPhone → brand

5) Summarization and VisuaLization Module (SVLM):

In the last process, the system will provide the overview of sentiment polarity for each product feature that people post though the visualization of feature-based summaries of opinions. We adopt three different visualization techniques as follows.
a) Radar Chart – to compare many properties of one interesting issue in one graph. Obviously, the greater the area covered by the graph, the greater the overall value.
b) Bar Chart – to highlight separate quantities, especially the differences between quantities in several categories of data.
c) Line Chart – to show the pattern how a value changes. Moreover, we can show multiple pieces of data by using multiple lines in the same graph.

4 Empirical Evaluations

4.1 Data set

In this study, the experiment is performed on the domain of smartphones since there are enormous number of conversations and thoughts related to smartphones on Twitter. Smartphone is a mobile telephone with extensive properties, such as a personal digital assistant, Internet browser and other applications. Once the technology progress has increased smartphone capabilities at an affordable price, the global smartphone market is incessantly growing. Smartphone becomes a part of many people daily life. As a result, tweets related to how people talk about smartphone are captured in our experiments. All tweets from public timeline spanning from March 2010 to June 2010 are collected via Twitter Search API and initially filtered using the following criteria.

(1) Tweets must contain related keywords associated with smartphone brands as shown in Table 5.
(2) Duplicated contents are ignored.
(3) Non-English tweets are also filtered out.
(4) Tweets that contain more than one product will not be considered because it may be a comparative sentence which is beyond our scope.

Operating System	Brand	Keywords
Apple iOS	iPhone	Iphone
Google Android	Motorola, HTC, Sumsung, Google Nexus	motorola milestone, motorola droid, motorola cliq, motorola devour, motorola backflip, htc, samsung galaxy, samsung omnia, samsung wave, nexus one
RIM Black Berry OS	BlackBerry	bb bold, bb torch, bb curve, blackberry
Symbian OS	Nokia	nokia n, nokia e, nokia c
Palm OS	Palm	palm pre, palm pixi

Table 5: Related terms and keywords for smartphones

4.2 Processing

As illustrated in Figure 1, we collected about 1,000,000 tweets using the Message Collecting Module (MCM). However, we randomly selected 100,000 tweets for the experiment to reduce processing time and to ensure the evenly distribution of each brand along the timeline. The tweet collection is separated into 10 groups equally for each collecting time period. So, 10,000 tweets are randomly selected from each group to put into the Opinion Filtering Module (OFM). To construct the classification model, we manually annotated 600 randomly selected tweets, half of which contain opinion and non-opinion tweets. Based on our previous study, we found that for opinion filtering model the Information Gain (IG) feature selection and the classification algorithm Support Vector Machines (SVM) yielded the best performance [Chamlertwat, 11]. After processing 100,000 tweets with OFM, we got the results of about 20,000 opinion tweets. Then, these tweets are sent to the Polarity Detecting Module (PDM) to analyze and assign the polarity scores for each tweet. The Feature Classification Module (FCM) will further extract product feature and brand mentioned in each tweet. Finally, the results will be displayed by the Summarization and VisuaLization Module (SVLM).

4.3 Results

The experimental results in Table 6 show that during the study period iPhone is the brand that people talk about the most, followed by Blackberry and HTC, respectively. However, after PDM processing, the rank is changed a little to iPhone, Blackberry, and Nexus. We rarely found conversation about Palm and Nokia in micro-blogs.

	MCM	OFM		PDM	
Blackberry	100,000	4,105	20.63%	548	13.72%
HTC		2,864	14.39%	428	10.71%
iPhone		8,091	40.65%	2,231	55.84%
Motorola		990	4.97%	157	3.93%
Nexus		2,305	11.58%	458	11.46%
Nokia		370	1.86%	48	1.20%
Palm		73	0.37%	16	0.40%
Samsung		1,105	5.55%	109	2.73%
Total	100,000	19,903		3,995	

Table 6: Number and percentage of the twitter posts

As illustrated in Table 7, people prefer to talk about the features, Application, Camera, and Power. Surprisingly, the number of complimented tweets outnumbers the number of complaints. For Blackberry, both positive and negative tweets for each feature are almost equally balanced; except the positive tweets for application is more than negative ones. For HTC, the number of positives outperforms the negatives in every feature. For iPhone, the favourite features are Application, Network, and

Camera. However, they have some concern about Accessories. For Motorola, the positive tweets are on Camera and Power. Similar to HTC, Nexus has good perception for every feature except Power. For Nokia, it is strong for their Network. For Palm, there is too little data to say whether it is good or bad. Finally, for Samsung, it is very outstanding for its Camera.

	Blackberry		HTC		iPhone		Motorola	
	+	-	+	-	+	-	+	-
Screen	1	0	1	0	4	1	0	0
Application	150	106	43	20	1071	473	3	3
Network	20	21	52	15	104	69	18	12
System	0	0	21	1	7	1	12	15
Camera	66	60	138	25	176	73	36	7
Capacity	2	4	1	1	4	3	0	1
Power	53	51	65	39	101	123	35	12
Sensor	1	0	0	0	0	0	0	0
Accessories	9	2	3	0	6	14	2	1
Size	2	0	3	0	0	0	0	0
	Nexus		**Nokia**		**Palm**		**Samsung**	
	+	-	+	-	+	-	+	-
Screen	0	1	0	0	0	0	7	0
Application	37	12	2	4	7	1	10	10
Network	33	14	17	5	1	0	3	2
System	4	0	0	0	0	0	0	0
Camera	175	139	6	3	2	0	62	7
Capacity	1	0	1	1	0	0	0	1
Power	16	19	4	0	4	0	4	1
Sensor	1	0	0	0	0	0	0	0
Accessories	1	3	5	0	1	0	0	2
Size	0	2	0	0	0	0	0	0

Table 7: Number of sentiment tweets for each brand from FCM

To summarize the result, the process of SVLM will visualize these numbers by three different techniques.

Radar chart – to compare each feature in a brand as illustrated in Figure 2. For example, we are considering HTC in 10 product aspects. Camera is obviously the best one. Power might be a majorly concerned feature since there are both positive and negative opinions in an interesting portion.

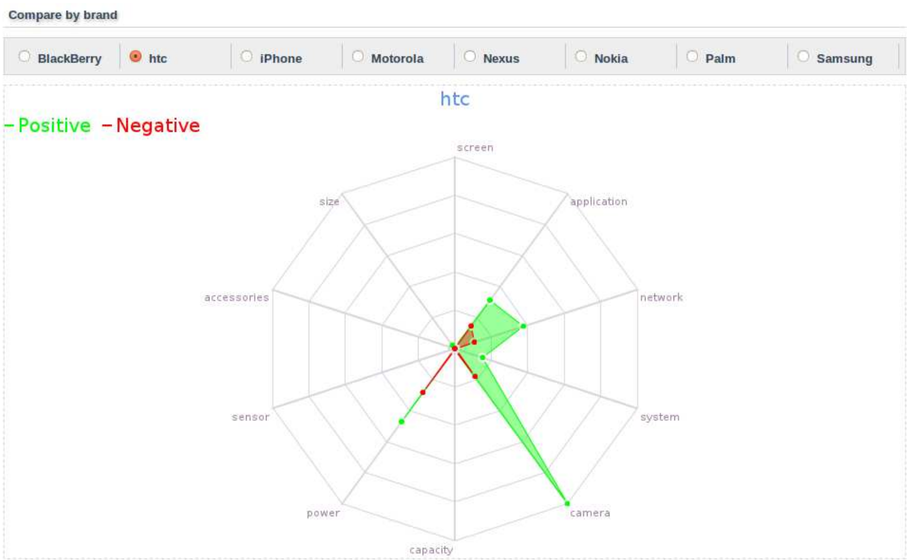

Figure 2: The comparisons of features for each brand

Bar chart – to compare each brand for the same feature as illustrated in Figure 3. For example, we are considering network feature. iPhone is the most talked about one while there are both high positive and high negative amount of opinions. HTC is quite impressive since there are double to three time positive opinions more than negative opinions.

Figure 3: The comparisons of brands for each feature

Line chart – to compare selected brand in every features as illustrated in Figure 4. For example, we choose to compare Blackberry with Motorola and Nexus.

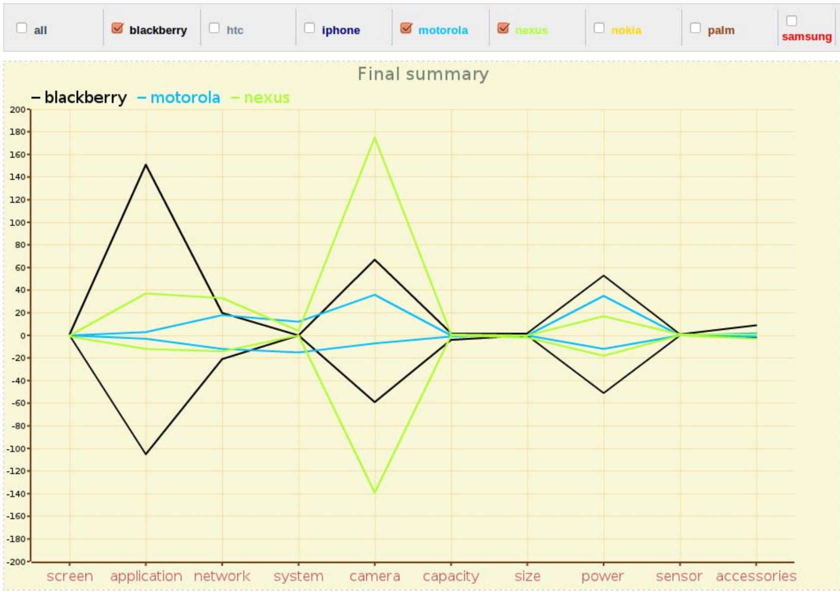

Figure 4: The comparisons of brands and features

4.4 Validation

We validate the system results asking three experts in smartphone industry to assign the scores for each smartphone brand. The score is based on the Likert scale from 1 to 5, where 5 is strongly agree and 1 is strongly disagree. After the experts understood our experiment and viewed the system results, we ask them to rate their confidence on the system results. The confidence scores are shown in Table 8.

Brand	**Confidence score from expert**					**Averag e weight**
	Strongly agree (5)	**Agree (4)**	**Undecided (3)**	**Disagree (2)**	**Strongly disagree (1)**	
Blackberry	0	1	2	0	0	3.33
HTC	1	1	1	0	0	4
iPhone	2	1	0	0	0	4.67
Motorola	0	2	1	0	0	3.67
Nexus	1	2	0	0	0	4.33
Nokia	0	0	2	1	0	2.67
Palm	0	0	2	1	0	2.67
Samsung	0	0	3	0	0	3
Average	**0.5**	**0.875**	**1.375**	**0.25**	**0**	

Table 8: Confidence scores from experts for each brand

As shown in Table 8, all experts majorly comply with the system results. The top-3 ranked scores fall into *Undecided*, *Agree* and *Strongly agree*, respectively. The top-3 brands, which received highest confidence scores, are iPhone, Nexus and Blackberry. These brands are ranked among the brands with highest gross sale in the market. The brands, which receive low confidence scores, are Nokia and Palm. Both receive only 2.67 weight score, which is under average. The reason may be due to the number of tweets for Nokia and Palm is relatively low compared to other brands. The low number of tweets directly reflects the low popularity of the brands. Moreover, there are some concerns, for example: not all HTC and Samsung models have good camera feature. About the network feature, it is not direct property of smartphone but it also depends on network operators.

5 Discussion

In this experiment, the system was performed on the combination of both lexicon-based and machine learning based approaches. Like Pang's work [Pang, 02], we use machine learning approach in our Opinion Filtering Model (OFM) to get high recall while SentiWordNet which based on the lexicon-based approach, using in the Polarity Detecting Module (PDM), can deliver high precision [Baccianella, 10].

Although there has been some previous research in sentiment analysis on micro-blog, no research has emphasized on the utilization of the end result. The result from Twitter Sentiment [Go, 09] provides only overview of consumer sentiment on a product, but it cannot specify the sentiment on any exact feature.

However, the result from the MSAS may not yield direct feedback for the product feature itself. For example, the result of iPhone shows that Apple might have some problem with its accessories, but the fact is most accessories are not manufactured by Apple. Apple allows third-party companies to build accessories for their iPhone. Therefore, the results from this system must have to comply with the real world situation first.

6 Conclusion

In this paper, we reported an exploratory study using our *Micro-blog Sentiment Analysis System* (MSAS) to discover consumer insight. Our work reconfirms that sentiment analysis on micro-blog, especially Twitter, can provide *supportive* information for producers in smartphone industry to make some decision about their next generation product. Our system, the MSAS, can gather information regarding product feature review without disturbing consumers, and the result is acceptable by experts in the field. Finally, we conclude that sentiment analysis on micro-blog is very useful tool for the consumer research, especially in the industries that customers spend their time on social media. The less data we can collect from social media, the more error we will suffer from the analyzed result.

7 Future works

Our future work is to improve the quality of the analyzed results. More research on algorithm will be performed to improve the performance and accuracy of the system. We would also study some solutions to handling some interesting and challenging issues, such as monitoring the opinions and sentiment from each single user over time. Other issues are to improve the algorithm in order to (1) handle tweets containing multiple features and multiple polarities in the same post, and (2) to handle multiple negations in the same post.

Acknowledgements

This research is supported by The 90th anniversary of Chulalongkorn University Fund (Ratchadaphiseksomphot Endowment Fund).

References

[Agarwal, 11] Agarwal, A., Xie, B., Vovsha, I., Rambow, O., and Passonneau, R.: Sentiment analysis of Twitter data. In Proceedings of the Workshop on Languages in Social Media (LSM '11). Association for Computational Linguistics, pp. 30-38, 2011.

[Baccianella, 10] Baccianella, S., Esuli, A., and Sebastiani, F.: SENTIWORDNET 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. In Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10), 2010.

[Bobo, 09] Bobo, L. D., and Dawson, M. C.: A change has come: Race, politics, and the path to the Obama presidency. Du Bois Review: Social Science Research on Race 6(1): pp. 1-14, 2009.

[Cashmore, 10] Cashmore P.: Facebook Founder on Privacy: Public Is the New "Social Norm" January 11, 2010 Retrieved from: http://mashable.com/2010/01/10/facebook-founder-on-privacy/

[Chamlertwat, 11] Chamlertwat, W., Bhattarakosol, P., and Haruechaiyasak, C.: Improving Sentiment Analysis on Twitter with Intention Classification. In Proceedings of 2011 3rd International Conference on Computer Engineering and Applications (ICCEA 2011), 2011.

[Esuli, 07a] Esuli, A., and Sebastiani, F.: PageRanking WordNet Synsets: An Application to Opinion Mining. In Proceedings of ACL-07, the 45th Annual Meeting of the Association of Computational Linguistics, pp. 424-431, 2007.

[Esuli, 07b] Esuli, A., and Sebastiani, F.: SentiWordNet: A Publicly Available Lexical Resource for Opinion Mining. In Proceedings of the 5th Conference on Language Resources and Evaluation, pp. 417-422, 2007.

[Ferreira, 11] Ferreira, A. : Android OS changes smartphone life cycle. February 16, 2011 Retrieved from http://www.theusdvista.com/business/android-os-changes-smartphone-life-cycle-1.2000033

[Glance, 05] Glance, N., Hurst, M., Nigam, K., Siegler, M., Stockton, R., and Tomokiyo, T.: Deriving marketing intelligence from online discussion. In Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining (KDD '05). ACM, pp. 419-428, 2007.

[Go, 09] Go, A., Bhayani, R., and Huang, L.: Twitter sentiment classification using distant supervision. Stanford University, Palo Alto, CA, USA, 2009.

[Hu, 04] Hu, M., and Liu, B.: Mining and summarizing customer reviews. In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '04). ACM, pp. 168-177, 2004.

[Jung, 08] Jung, J.J.: Taxonomy alignment for interoperability between heterogeneous virtual organizations. Expert Systems with Applications 36(4): 2721-2731, 2008.

[Jung, 09] Jung, J.J.: Contextualized Query Sampling to Discover Semantic Resource Descriptions on the Web. Information Processing & Management 45(2): pp. 283-290, 2009.

[Jung, 10a] Jung, J.J.: Ontology Mapping Composition for Query Transformation on Distributed Environments. Expert Systems with Applications 37(12): pp. 8401-8405, 2010.

[Jung, 10b] Jung, J.J.: Integrating Social Networks for Context Fusion in Mobile Service Platforms. Journal of Universal Computer Science 16(15): pp. 2099-2110, 2010.

[Jung, 11a] Jung, J.J.: Boosting Social Collaborations Based on Contextual Synchronization: An Empirical Study. Expert Systems with Applications 38(5): pp. 4809-4815, 2011.

[Jung, 11b] Jung, J.J.: Service Chain-based Business Alliance Formation in Service-oriented Architecture. Expert Systems with Applications 38(3): pp. 2206-2211, 2011.

[Jung, 12] Jung, J.J.: Evolutionary Approach for Semantic-based Query Sampling in Large-scale Information Sources. Information Sciences 182(1): pp. 30-39, 2012.

[Kouloumpis, 11] Kouloumpis, E., Wilson, T., and Moore, J.: Twitter Sentiment Analysis: The Good the Bad and the OMG!. The 5th International AAAI Conference on Weblogs and Social Media (ICWSM-11), 2011.

[Lake, 11] Twitter Sentiment Analysis. Western Michigan University, Kalamazoo, MI, USA, 2011.

[Liu, 05] Liu, B., Hu, M., and Cheng, J.: Opinion observer: analyzing and comparing opinions on the Web. In Proceedings of the 14th international conference on World Wide Web (WWW '05). ACM, pp. 342-351, 2005.

[Liu, 11] Liu, B.: Opinion Mining and Sentiment Analysis, WEB DATA MINING, Data-Centric Systems and Applications, Part 2, pp. 459-526, 2011.

[Pang, 02] Pang, B., Lee, L., and Vaithyanathan, S.: Thumbs up?: sentiment classification using machine learning techniques. In Proceedings of the ACL-02 conference on Empirical methods in natural language processing - Volume 10 (EMNLP '02), pp. 79-86, 2002.

[Stone, 08] Stone, M., Bond, A., and Foss, B.: CONSUMER INSIGHT How to use data and market research to get closer to your customer: Kogan Page, 2008.

[Turney, 02] Turney, P.: Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In Proceedings of the 40th Annual Meeting on Association for Computational Linguistics (ACL '02), pp. 417-424, 2002

[Twitter blog, 11] Your world, more connected. August 01, 2011 Retrieved from http://blog.twitter.com/2011/08/your-world-more-connected.html

[Twitter website, 11] Using the Twitter Search API. October 10, 2011 Retrieved from https://dev.twitter.com/docs/using-search

[Wilson, 05] Wilson, T., Wiebe, J., and Hoffmann, P.: Recognizing contextual polarity in phrase-level sentiment analysis. In Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing (HLT '05), pp. 347-354, 2005.

[Zabin, 08] Zabin, J., Jefferies, A.: Social Media Monitoring and Analysis: Generating Consumer Insights from Online Conversation. Aberdeen Group Benchmark Report, 2008.

[Zhai, 11] Zhai, Z., Liu, B., Xu, H., and Jia, P.: Clustering product features for opinion mining. In Proceedings of the fourth ACM international conference on Web search and data mining (WSDM '11). ACM, pp. 347-354, 2011.

[Zhang, 11] Zhang, L., Ghosh, R., Dekhil, M., Hsu, M., and Liu, B.: Combining Lexicon-based and Learning-based Methods for Twitter Sentiment Analysis. HP Laboratories, HPL-2011-89, 2011.

[Zhe, 10] Zhe, X.: A Sentiment Analysis Model Integrating Multiple Algorithms and Diverse Features. Ohio State University, Columbus, OH, USA, 2010.