Review-based measurement of customer satisfaction in mobile service: Sentiment analysis and VIKOR approach

Article *in* Expert Systems with Applications · March 2014

DOI: 10.1016/j.eswa.2013.07.101

Daekook Kang
Inje University

Expert Systems with Applications 41 (2014) 1041–1050

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Review-based measurement of customer satisfaction in mobile service: Sentiment analysis and VIKOR approach

Daekook Kang [1], Yongtae Park *

Department of Industrial Engineering, School of Engineering, Seoul National University, San 56-1, Shilim-Dong, Kwanak-Gu, Seoul 151-742, Republic of Korea

A R T I C L E I N F O

Keywords:
Customer satisfaction
Mobile service
Customer review
Sentiment analysis
VIKOR

A B S T R A C T

With the rapid growth and dissemination of mobile services, enhancement of customer satisfaction has emerged as a core issue. Customer reviews are recognized as fruitful information sources for monitoring and enhancing customer satisfaction levels, particularly as they convey the real voices of actual customers expressing relatively unambiguous opinions. As a methodological means of customer review analysis, sentiment analysis has come to the fore. Although several sentiment analysis approaches have proposed extraction of the emotional information from customer reviews, however, a lacuna remains as to how to effectively analyze customer reviews for the purpose of monitoring customer satisfaction with mobile services. In response, the present study developed a new framework for measurement of customer satisfaction for mobile services by combining VIKOR (in Serbian: *ViseKriterijumsa Optimizacija I Kompromisno Resenje*) and sentiment analysis. With VIKOR, which is a compromise ranking method of the multicriteria decision making (MCDM) approach, customer satisfaction for mobile services can be accurately measured by a sentiment-analysis scheme that simultaneously considers *maximum group utility* and *individual regret*. The suggested framework consists mainly of two stages: data collection and preprocessing, and measurement of customer satisfaction. In the first, data collection and preprocessing stage, text mining is utilized to compile customer-review-based dictionaries of attributes and sentiment words. Then, using sentiment analysis, sentiment scores for attributes are calculated for each mobile service. In the second stage, levels of customer satisfaction are measured using VIKOR. For the purpose of illustration, an empirical case study was conducted on customer reviews of mobile application services. We believe that the proposed customer-review-based approach not only saves time and effort in measuring customer satisfaction, but also captures the real voices of customers.

1. Introduction

With the recent advances in wireless technology, the number of people using mobile devices has increased, and the development of mobile services, already rapid, has accelerated (Wang, Lin, & Luarn, 2006). And as mobile services proliferated in number, customers began to develop and express opinions on services they had experienced. Compared with the e-business environment, this phenomenon has occurred more often, because people have the ability to access information anytime and anywhere, as the move towards use of mobile technologies rapidly accelerates (Perry, O'Hara, Sellen, Brown, & Harper, 2001). As a result, the necessity of mobile service providers enhancing and maintaining high levels of customer satisfaction by monitoring and listening to customer voices has emerged as a core issue (Choi, Seol, Lee, Cho, & Park, 2008; Deng, Lu, Wei, & Zhang, 2010; Kim, Park, & Jeong, 2004; Kim & Yoon, 2004; Kuo, Wu, & Deng, 2009; Lee, Lee, & Feick, 2001; Wang & Lo, 2002; Woo & Fock, 1999).

Customer review is utilized frequently to monitor and, so, enhance customer satisfaction with mobile services. Customer review is an attractive alternative to average ratings, in that the latter, given their typically bimodal distribution, often do not provide particularly helpful information on customer satisfaction (Abulaish, Jahiruddin, & Doja, 2009). In bimodal distributions, average ratings are either extremely high or extremely low, and as such, might not convey very much information that is helpful for the purposes of monitoring and enhancing customer satisfaction. Actual customer reviews, by contrast, include customers' voices, which are starkly revealing as to the positive and negative aspects of a given service (Zhang, Narayanan, & Choudhary, 2010). Thus, actual customer reviews have been recognized as a valuable information source for monitoring and enhancing customer satisfaction with mobile services.

Mobile service customer review has characteristics distinct from those of other types of e-business customer review. For

* Corresponding author. Tel.: +82 2 880 8358; fax: +82 2 889 8560.
E-mail addresses: daekooky@snu.ac.kr (D. Kang), parkyt1@snu.ac.kr, parkyt@cybernet.snu.ac.kr (Y. Park).
[1] Tel./fax: +82 2 878 3511.

0957-4174/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2013.07.101

example, unlike web service customer reviews, in mobile services, each customer review has a concise core content focused on important criteria. This is due to the fact that many people generally write their opinions through their mobile devices: it is relatively more difficult to write review contents with a mobile device than with a keyboard, and so, in the mobile environment, people tend to write their reviews brief and to the point with regard to the important key criteria. Moreover, customer reviews of mobile services tend to be relatively clear and unambiguous.

For mobile service providers, customer reviews can be a useful and relatively low-effort means of understanding the minds of customers. Knowledge thus gained, furthermore, is quite reliable, since customers write reviews voluntarily. Customer reviews disseminate information such as the functional characteristics, advantages and disadvantages of certain types of mobile service. However, mobile service providers efforts at monitoring customer reviews have suffered due to the lack of an effective methodology for effective analysis. Also, from the customer's point of view, many people only consider information about overall ratings when they first select a specific alternative (Poston & Speier, 2005), due to the difficulty of reading a huge number of customer reviews to obtain information on all of the many alternative-mobile-service criteria (Hu & Liu, 2004a,b).

Certainly, there is a great need for effective analysis of customer reviews. For maximization of the value of information obtained, customer satisfaction with mobile services needs to be analyzed by way of simultaneous consideration of customers' heterogeneous preferences for multiple and various service attributes, rather than on the basis, simply, of general, overall ratings of a whole service.

In order to obtain information on customer satisfaction with mobile services with respect to each service attribute, sentiment analysis can be utilized. Sentiment analysis is a technique for identifying the ways in which sentiments are expressed in text and for determining whether they represent positive or negative feelings toward a specific product or service (Nasukawa & Yi, 2003). Sentiment analysis has several specific advantages. First, it can extract emotional information from text semi-automatically. Based on predefined sentiment dictionaries, sentiment analysis quickly captures the polarity of a specific attribute by analyzing vast numbers of customer reviews. This entails transformation of textual information into a single number or word revelatory of a certain level of customer satisfaction. Second, sentiment analysis provides customer satisfaction information in terms of the several and specific attributes of a service.

However, for the most effective decision-making support, overall customer satisfaction with a whole service, not just with individual attributes, also needs to be provided. Sentiment analysis is less than suitable for this purpose, because, first, the multiple attributes of a mobile service affect customer satisfaction simultaneously, but their respective significances differ; and, second, attributes impacting on customer satisfaction differ greatly among the various categories. Measurement of overall customer satisfaction at the whole-service level, therefore, calls for an alternative methodology. The multicriteria decision making (MCDM) approach is pertinent for reflecting these points because its strength is the simultaneous consideration of various criteria. It can also reflect the importance of each criterion in evaluating alternatives.

Among the various MCDM approaches, VIKOR (in Serbian: *Vise-Kriterijumsa Optimizacija I Kompromisno Resenje*) was utilized in the present study. VIKOR is a compromise ranking method to optimize the multi-response process (Opricovic, 1998). It is used for ranking and selecting from a set of alternatives in the presence of conflicting criteria based on the closeness to the ideal solution. Compared with the other MCDM approaches, VIKOR has some advantages. First, with respect to mobile services regarding which customer reviews are frequently generated, unsatisfactory attributes affect the selection of the whole mobile service remarkably, because these kinds of mobile service have various similar services in the same categories. For example, in App Store, Apple's market of mobile application services and a forum for numerous customer reviews, many similar services exist. In the case of *music player*, there are more than 1500 mobile application services. In such a situation, when a customer chooses one service he prefers, he rules out, in an initial, screening step, others that have unsatisfactory attributes. And this step, in fact, is very essential to mobile service selection. Here, VIKOR is advantageous in that it provides not only information on *maximum group utility*, which is the utility that reflects consideration of all relevant criteria, but also information on *individual regret*, which is the marker of consideration of the most unsatisfactory criterion. Second, mobile services are classified into many categories, the characteristics of which are distinguishable. Thus, according to the mobile service category, the evaluator can set the weights of *maximum group utility* and *individual regret* differently. For these reasons, VIKOR, compared with other MCDM approaches, can be better applied for sentiment-analysis-result-based measurement of total customer satisfaction with mobile services.

Accordingly, this study presents a new framework for measuring customer satisfaction with mobile services that uses sentiment analysis and VIKOR to analyze customer reviews. The proposed framework consists mainly of two stages: *data collection and preprocessing* and *measurement of customer satisfaction.* In the *data collection and preprocessing* stage, text mining is used to compile dictionaries of attributes and sentiment words based on the review data. Then, using sentiment analysis, sentiment scores of attributes are calculated in terms of each mobile service. In the *measurement of customer satisfaction* stage, a sentiment score matrix is constructed, after which, using VIKOR, total customer satisfaction is calculated in consideration of the various attributes. In this study, an empirical case study was conducted based on the customer reviews of mobile application services.

The remainder of this paper is organized as follows. Section 2 introduces the basic concepts along with definitions of sentiment analysis and VIKOR. Section 3 proposes the overall measurement framework and explains in detail the process steps therein. Section 4 presents a case study in which the proposed framework and approach were applied, and explains the results. Finally, Section 5 concludes the paper and anticipates future research.

2. Literature review

2.1. Sentiment analysis

Sentiment analysis is a method for identifying the ways in which sentiment is expressed in texts and whether such expressions include positive or negative opinions on a certain product or service (Nasukawa & Yi, 2003). Specifically, it is an analysis of the opinions and emotions contained in direction-based text. Sentiment analysis typically has been employed at two levels (Wilson, Wiebe, & Hoffmann, 2005). First, sentiment analysis is conducted at the document level to distinguish positive from negative reviews (Beineke, Hastie, & Vaithyanathan, 2004; Pang & Lee, 2004; Pang, Lee, & Vaithyanathan, 2002; Turney, 2002). Second, it has been conducted at the sentiment level or phrase level to perform tasks such as multi-perspective question answering and summarization, opinion-oriented information extraction, and customer review mining (Hu & Liu, 2004a,b; Kamp and Mark, 2002; Liu, Hu, & Cheng 2005; McDonald, K., & T., 2007; Popescu & Etzioni, 2005; Täckström & McDonald, 2011; Wilson, Wiebe, & Hoffmann, 2009; Wilson et al., 2005; Zhang, Xu, & Wan, 2008).

Generally, sentiment analysis begins with sentiment expression with regard to a given object, and then distinguishes a lexicon of positive and negative words and phrases (Kang & Park, 2012). There are three types of lexicons: positive polarity (e.g. *wonderful*), negative polarity (e.g. *bad*, *pessimistic*, *terrible*), and contextual polarity (i.e. phrases in which a word or words can carry different meanings in different contexts).

There has been a steady academic interest in sentiment analysis. The body of research usually has focused on document-level classification of overall sentiment that distinguishes positive from negative reviews. Turney (2002) proposed a simple algorithm for classification of reviews as *recommended* or *not recommended* based on the value of the average semantic orientation of phrases containing adjectives or adverbs. Pang and Lee (2004) presented a novel machine-learning method that applies text-categorization techniques to the subjective portions of a document based on minimum cuts. Beineke et al. (2004) extended the traditional sentiment classification procedure by re-interpreting it as a Naïve Bayes model.

Additionally, many studies have employed advanced techniques to analyze sentiment at the sentence or phrase level. For example, Kamps and Marx (2002) explored how the structure of the Word-Net lexicon database might be used to assess affective or emotive meaning, specifically by formulating measures based on Osgood's semantic differential technique. Also, Wilson et al. (2005) presented a new approach to phrase-level sentiment analysis that classifies whether an expression is neutral or polar and, subsequently, disambiguates the polar expressions. Meanwhile, McDonald et al. (2007) constructed a model for joint classification of sentiment based on standard sequence classification techniques that utilize a constrained Viterbi algorithm, which is a dynamic programming algorithm for finding the most likely sequence of hidden states. Recently, Täckström and McDonald (2011) introduced two variants of a semi-supervised latent variable model for sentence-level sentiment analysis.

Although the aforementioned research dealt broadly with sentiment analysis at both the document and sentence levels, it focused only on the techniques themselves, and thus is of limited usefulness for deriving useful information to measure customer satisfaction. Accordingly, there is a need to obtain additional information on customer satisfaction using the results derived from sentiment analysis. Accordingly, this paper proposes a sentiment-analysis-based framework for measuring customer satisfaction.

2.2. VIKOR

Multicriteria decision making (MCDM) is one of the most prevalent methods for resolving conflict management issues (Deng & Chan, 2011). MCDM deals with decision and planning problems by consideration of multiple criteria and the importance of each (Haleh & Hamidi, 2011). Among the many MCDM methods, VIKOR is a compromise ranking method to optimize the multi-response process (Opricovic, 1998). It uses a multicriteria ranking index derived by comparing the closeness of each criterion to the ideal alternative. The core concept of VIKOR is the focus on ranking and selecting from a set of alternatives in the presence of conflicting criteria (Opricovic, 2011). In VIKOR, the ranking index is derived by considering both the *maximum group utility* and minimum *individual regret* of the opponent (Liou, Tsai, Lin, & Tzeng, 2011).

VIKOR denotes the various n alternatives as $a_1, a_2, \ldots, a_n$. For an alternative a_i, the merit of the jth aspect is represented by f_{ij}; that is, f_{ij} is the value of the jth criterion function for the alternative a_i, n being the number of criteria. The VIKOR procedure is divided into the following five steps:

(1) Determine the best f_j^* and worst f_j^- values of all criterion functions. If the jth criterion function represents a merit, then

$$f_j^* = max_i f_{ij}, \quad f_j^- = min_i f_{ij} \tag{1}$$

(2) Compute the values S_i and R_i, $i = 1,2,3,\ldots,m$, by the relations

$$S_i = \sum_{j=1}^{n} \frac{w_j\left(f_j^* - f_{ij}\right)}{f_j^* - f_j^-} \tag{2}$$

$$R_i = max\left[\frac{w_j\left(f_j^* - f_{ij}\right)}{f_j^* - f_j^-}\right] \tag{3}$$

where w_j is the weight of the jth criterion which expresses their relative importance of the criteria.

(3) Compute the value Q_i, $i = 1,2,3,\ldots,m$, by the relation

$$Q_i = v\left[\frac{S_i - S^*}{S^- - S^*}\right] + (1 - v)\left[\frac{R_i - R^*}{R^- - R^*}\right], \tag{4}$$

where $S^* = \min_i S_i$, $S^- = \max_i S_i$, $R^* = \min_i R_i$, $R^- = \max_1 R_1$ and v is the weight of the strategy of *maximum group utility*, whereas $(1 - v)$ is the weight of the *individual regret*. Here, when v is larger than 0.5, the index of Q_i follows majority rule

(4) Rank the alternatives, sorting by the values S, R and Q, in decreasing order.

VIKOR is advantageous in the context of multicriteria-based decision making, particularly in situations where the decision maker is not able, or does not know how he/she expresses his/her preference at the early stage of system design. Because VIKOR provides a maximum "group utility of majority" represented by min S, and a minimum "*individual regret* of opponent" represented by min R, decision makers can determine compromise solutions based on their negotiated preferences. Thus, where unsatisfactory attributes can remarkably affect the selection of an entire service, VIKOR, compared with other MCDM methods, is useful for achieving the purpose of this study, because the decision maker can set the weights of *maximum group utility* and *individual regret* differently according to various situations.

3. Proposed approach

3.1. Overview

The framework of proposed approach is portrayed in Fig. 1. As shown, the process mainly consists of two parts with respective to overall perspective: *data collection and preprocessing* and *measurement of customer satisfaction*. In the *data collection and preprocessing* stage, first, customer reviews on the web are collected. Then, on the basis of those data, text mining is used to compile dictionaries of attributes and sentiment words, after which the review data is transformed, with reference to the dictionaries, into keyword vectors of customers' opinions. A keyword vector is comprised of the sentiment rating scores for each important attribute. In the *measurement of customer satisfaction* stage, a matrix of customer satisfaction is constructed by integrating the keyword vectors according to the various mobile services, and the weights of each attribute are calculated. Lastly, the final levels of customer satisfaction for a given mobile service are evaluated by considering customer satisfaction with the attributes and the weights of those attributes simultaneously with VIKOR.

3.2. Preprocessing data

In the preprocessing phase, mobile service review data is collected from a relevant website. Generally, customer reviews include subjective evaluations of attributes that are components of a whole service. Those subjective evaluations use polarity scores

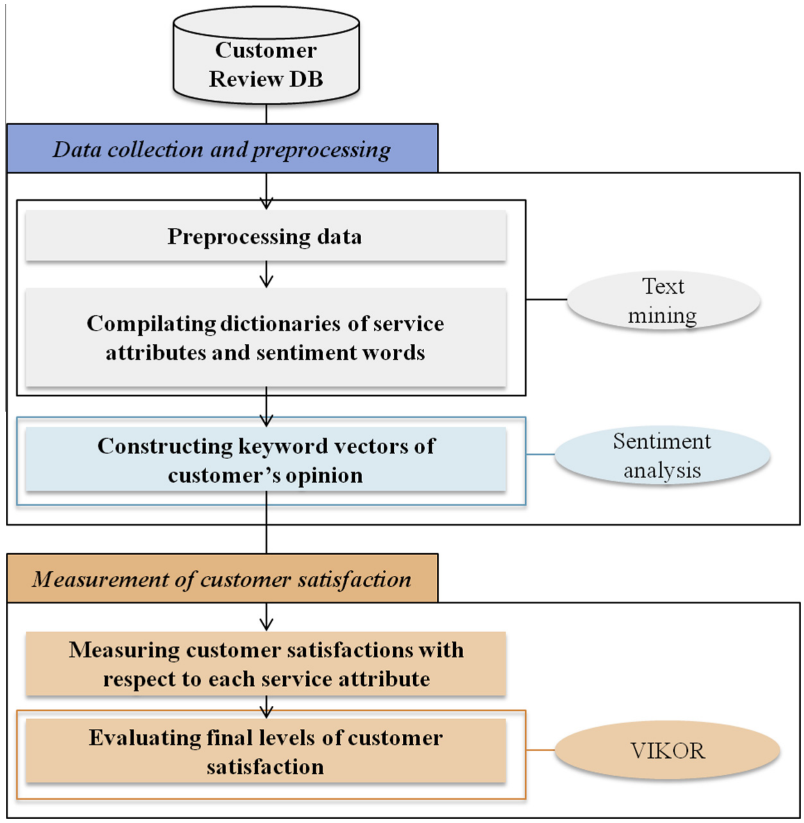

Fig. 1. Overall research framework.

to represent the merits or demerits of attributes. According to the presence of attribute mentions (i.e. the "*term presence*") in customer reviews, the attributes that customers consider to be important can be identified.

In this study, review data on similar mobile services were collected. Since similar mobile services generally belong to the same category, the important attributes that customers consider do no greatly differ among services. For this reason, common attributes can be extracted from review data of similar mobile services and analyzed to measure and compare their respective customer satisfactions.

3.3. Compilation of dictionaries of attributes and sentiment words

In this phase, two dictionaries are compiled based on the collected review data: a dictionary of attributes, and a dictionary of sentiment words. The dictionary of attributes comprises the important attributes among the various services, which attributes are extracted by text-mining. In the present study, extraction was carried out using *Stanford parser*, a program that determines the grammatical structures of sentences. This program assigns Parts-Of-Speech (POS) tags to all words based on the contexts in which they appear (Abulaish et al., 2009). This POS information is utilized to locate different types of information of interest in text documents. More specifically, it provides information on which groups of words go together (as "phrases") and which words are subjects or objects of verbs. Generally, attributes in the review data are expressed as noun phrases, and opinions are expressed as adjectives. In the present study, noun phrases were extracted from the review data using *Stanford parser*, from which phrases important attributes were derived by experts.

In the compilation of the dictionary of sentiment words, sentiment words, typically expressed in verb phrases, adjective phrases and adverbial phrases, were extracted, again using *Stanford parser*. Then, the polarities of the derived sentiment words were classified into several groups: positive polarity, negative polarity, and neutral polarity. For each sentiment word, the sentiment score was expressed with reference to WordNet (Miller, 1998), because WordNet is a large electronic lexical database for English and English nouns, verbs, and adjectives are well organized into synonym sets, each representing one underlying lexical concept. More specifically, the sentiment score was calculated as an integer number between -2 and 2 with reference to sentiment classification scheme of Moreo, Romero, Castro, and Zurita (2012). Moreo et al. (2012) classified sentiment expression into 5 categories; *very positive, positive, neutral, negative, and very negative*. Thus, in this study, if a sentiment word is strongly positive, its value is 2, and if positive, 1. Meanwhile, if a sentiment word is strongly negative, its value is -2, and if the polarity for a certain attribute is negative, its value is -1. Lastly, if there is no sentiment, the word is expressed as having a value of 0.

3.4. Construction of keyword vectors of customers' opinions

After compiling the two dictionaries, review documents are transformed into keyword vectors of customers' opinions. In review documents, if attribute and sentiment words appear in the same sentence, that sentence is considered to be a customer's opinion. In this case, the sentiment score is counted for the corresponding attribute. For example, suppose that a customer writes the following review of a specific service:

> "Graphics are perfect. Accuracy is also good. Graphics are good. Camera function is somewhat normal.But Sound is terrible. Also, Update should be necessary."

Then, the keyword vectors of that customer's opinions are constructed with reference to the dictionaries. In the present study, those keyword vectors were constructed as

$$\mathrm{CR}_i = (\ \mathrm{PA}_{i1},\ \mathrm{PA}_{i2}, \ldots,\ \mathrm{PA}_{im}) \tag{5}$$

CR_i: Review of customer *i*.
PA_{ij}: Customer *i*'s polarity for *j*th attribute.
Fig. 2 shows examples of construction process of keyword vectors based on dictionaries of attributes and sentiment words.

3.5. Measurement of customer satisfaction with respect to each attribute

Once the sentiments of attributes have been detected in individual customer review, it is desirable to extend this to every customer reviews that attributes appear in. One of the most straightforward, popular, and simple ways to accomplish this is a linear combination of all polarity from every documents (Mejova, 2009). The merit of this ways to apply the present study is that it assumes there is no difference between the importances of documents, because the importances of customer reviews are the same. Thus, in this study, the levels of customer satisfaction for each attribute are measured by summing up the polarity scores from customer reviews as follows

$$\mathrm{CS} = \sum_{i=1}^{n} \mathrm{CR}_i(\ \mathrm{PA}_1,\ \mathrm{PA}_2, \ldots,\ \mathrm{PA}_m) \tag{6}$$

CS: Customer satisfaction for each service.
PA_k: Sum of polarity scores of *n* customers for *k*th attribute.
Meanwhile, in order to measure the weights of attributes, weighting scheme needs to be established, as there are many ways to measure attributes' weights in opinion mining and sentiment analysis. Traditionally, term frequency – inverse document frequency (td-idf), which is a term frequency-based numerical statistic that reflects how important a word is to a document, has been well utilized for weighting attributes (Chi, Prolli, & Chen, 2001; Djoerd, 2000; Ponte & Croft, 1998; Salton & Buckley, 1988; Zhang, Yoshida, & Tang, 2011); but in contrast, recently, Pang et al. (2002) obtained better performance using *term presence* rather than term frequency. That is, binary-valued feature vectors in which the entities merely indicate whether a term occurs (value 1) or not (value 0) showed a more effective performance for review polarity classification than did real-valued feature vectors in which entity values increase with the occurrence frequency of the corresponding term. For this reason, in this study, *term presence* – based weighting scheme is adopted for measure the weights of attributes. The *term presence* and weights of attributes are formulated as

$$\vec{d} := (n_1(d), n_2(d), \ldots, n_m(d))$$

$$W = (w_1, w_2, \ldots, w_m)$$

$$w_k = \frac{n_k(d)}{\sum_{k=1}^{m} n_k(d)} \tag{7}$$

$n_k(d)$: Number of times *k*th attribute occurs in document $d w_k$: Weights of *k*th attributes
After summing up the polarity scores for customer reviews and measuring the weights of attributes, in the next step, the polarity scores for attributes are normalized to a number between 0 and 1 for calculation convenience, as shown in Fig. 3.

3.6. Evaluation of final level of customer satisfaction

In this phase, VIKOR is used to evaluate the final level of customer satisfaction by considering all attributes, as based on the derived decision matrix that includes their weighted polarity scores. In VIKOR, a service alternative is evaluated by measuring only its closeness to an ideal solution. The final level of customer satisfaction is determined by adjusting the opponent's weights of *maximum group utility* and minimum *individual regret*. According to the importances of *maximum group utility* and minimum *individual regret* assigned, respectively by the evaluator, their weights for the opponent can be set variously to reflect different situations.

4. Empirical case study

To illustrate the efficacy of the proposed approach, an empirical case study was conducted. Recently, given the proliferation of smartphones, the number of mobile application services has increased dramatically. In the case of App Store which is Apple's

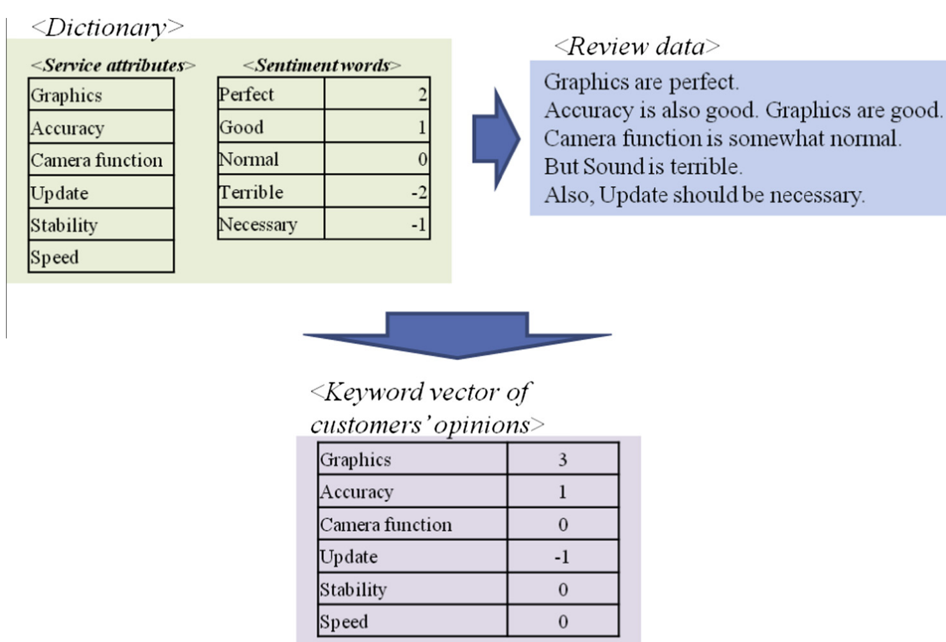

Fig. 2. Example of construction of keyword vector.

Service alternatives	Polarity scores of service attributes		
	Attribute 1	*Attribute 2*	*Attribute 3*
Service 1	-2	11	1
Service 2	5	4	-5
Service 3	-3	7	15
Term presence	20	50	30

Service alternatives	Normalized polarity scores of service attributes		
	Attribute 1	*Attribute 2*	*Attribute 3*
Service 1	0.15	0.80	0.30
Service 2	0.50	0.45	0.00
Service 3	0.10	0.60	1.00
Weight	0.2	0.5	0.3

Fig. 3. Examples of normalization of polarity scores of attributes.

mobile application market, the number of registered mobile application services is already more than one-half million. Since people can easily download a mobile application service with their mobile device, anytime, anywhere, many of them disingenuously write reviews of purchased services. Thus, a vast number of customer reviews has emerged in the online market. For this reason, the present empirical case study focused on customer reviews of mobile application services in App Store.

4.1. Preprocessing data

In this phase, App Store customer review data was preprocessed. In App Store, there are 20 categories, as among which are social networking, entertainment, and navigation. Social networking was selected for the study, owing to the fact that in this category, there are many similar services, each of which has similar attributes compared with the other service categories. The data source was *AppStoreHQ*, a website that provides reviews of mobile application services from blogs, Twitter and YouTube. Specifically, 1487 reviews were collected from 8 mobile application services in the social networking category, as shown in Table 1.

4.2. Compilation of dictionaries of attributes and sentiment words

The attributes customers considered important could be extracted by analysis of the collected customer reviews. First, *Stanford parser* was used to extract noun phrases. Then, the important attributes were derived based on the judgment of experts. In the result, 9 attributes were selected for measurement of customer satisfaction with mobile application services in the category of social networking services, as shown in Table 2.

Table 1
Number of customer reviews in each category of mobile application service.

Mobile application service	Number of customer reviews
Bump	166
Facebook	303
Foursquare	144
Fring	112
Google+	152
Skype	231
Twitter	271
WordPress	108

Second, the dictionary of sentiment words was compiled based on the words expressed in verb phrases, adjective phrases and adverbial phrases extracted using *Stanford parser*. In the result, a total of 287 sentiment words were derived, which were scored by experts as a number between -2 and 2, as shown in Fig. 4.

In this study, a large positive number indicated a highly positive sentiment toward a certain object. Conversely, a large negative number reflected a highly negative sentiment. For example, the word *best* was scored as 2, whereas *good* was scored as 1. Likewise, the word *terrible* was scored as -2, and the word *bad*, as -1.

4.3. Construction of keyword vectors of customers' opinions

Based on the two dictionaries, 1487 review documents were transformed into keyword vectors of customers' opinions. For the 9 attributes, sentiment scores were calculated using sentiment analysis. In this study, we created a C-language-based program, "*Sentiment Detecting*," to determine whether a given attribute and sentiment word appeared in the same sentence or not.

4.4. Measurement of customer satisfaction with respect to each attribute

First, keyword vectors of customers' opinions were formulated for 8 mobile application services, as shown in Table 3. Then, the attributes' polarity scores were normalized to a number between

Table 2
Criteria for measurement of customer satisfaction with mobile application service in category of social networking service.

Aspect and criterion	Description
Download (C_1)	Ease of downloading of various contents
Interface (C_2)	Accessibility of functions through simple operation
Message (C_3)	Smoothness of communication through message
Notification (C_4)	Immediacy of notification about new tidings
Photo (C_5)	Diversity of functions for uploading and decorating photos
Post (C_6)	Smoothness of uploading of text, photos, and music
Search (C_7)	Ease of searching of various information
Touch (C_8)	Ease of clicking and dragging of various contents
Update (C_9)	Continuous updates for bug fixing, improvement of functions, and addition of new functions

Sentiment word	Score	Sentiment word	Score	Sentiment word	Score	Sentiment word	Score	Sentiment word	Score	Sentiment word	Score	Sentiment word	Score	Sentiment word	Score
fickle	-1	super-high	2	frustrating	-2	clever	1	unfriend	-1	boring	-1	enjoyable	1	Audible	1
automatic	1	compelling	2	fast	1	strange	-1	unfriending	-1	invaluable	-1	skeptical	-1	first	1
little	-1	well-designed	1	interesting	1	awkward	-1	revolutionary	2	terrific	2	controversial	-1	new	1
old	-1	high-quality	2	pleasant	1	creative	1	meaningful	1	Fun	1	dirty	-1	worth	1
unaware	-1	powerful	2	cheap	1	fine	1	Improved	1	light	1	all-new	1	complete	1
clear	1	scant	-1	overwhelming	2	incredible	2	flexible	1	Lovely	1	alive	1	stable	1
available	1	desperate	-1	buggy	-1	positive	1	fantastical	2	tired	-1	optimized	1	malicious	-1
cool	1	adequate	1	disabled	-2	stupid	-1	wide-ranging	1	inappropriate	-1	autocomplete	1	top	2
ugly	-1	must-have	2	Wonderful	2	dazzling	2	lousy	-2	Hysterical	-1	slim	1	robust	1
fake	-1	broad	1	impressive	1	impossible	-1	unfinished	-1	obnoxious	-2	unusual	-1	believable	1
high	1	most-used	1	innovative	2	wise	1	superior	1	amusing	1	effortless	-1	keen-eyed	1
super	2	productive	1	inevitable	2	countless	-1	sweet	1	remarkable	1	intriguing	2	few	-1
easy	1	insightful	1	accurate	1	addictive	1	easy-to-use	1	fabulous	2	attractive	1	limited	-1
Priceless	2	surprising	2	ideal	1	welcome	1	Terrible	-2	artistic	1	new-style	1	quick	1
amazing	2	awesome	2	neat	1	number-one	1	weak	-1	distinct	1	disable	-1	beefy	-1
great	2	lazy	-1	exact	1	dedicated	1	meager	-1	whimsical	1	worst-case	-2	beautiful	1
time-sucking	-2	useful	1	bustling	-1	broken	-1	frivolous	-1	soooo	2	uncharted	-1	brand-new	1
fantastic	2	enough	1	worthy	1	snazzy	1	poor	-2	satisfied	1	error-free	1	sparkly	1
good	1	narwhal-loving	1	unique	1	unavailable	-1	unhappy	-1	glossy	1	likeable	1	narcissistic	-1
pioneering	1	excellent	2	interested	1	ridiculous	-1	difficult	-1	outstanding	2	happier	1	elegant	1
clean	1	exciting	2	surprised	1	silly	-1	puzzling	-1	dissapoint	-1	quicker	1	modern	1
full-fledged	1	awful	-1	unreal	-1	tasty	1	futuristic	1	unspecified	-1	slower	-1	uninterrupted	1
simple	1	vital	1	fascinating	2	proud	1	naughty	-1	cursory	-1	clearer	1	familiar	1
live	1	nice	1	active	1	long-running	1	messy	-2	tempting	1	greatest	2	exquisite	2
popular	1	sound	1	enhanced	1	top-nominated	2	severe	-2	unclear	-1	easiest	2	slow	-1
unsexy	-1	colorful	1	negative	-1	highly-coveted	2	unorganized	-1	sluggish	-1	coolest	2		

Fig. 4. Dictionary of sentiment words.

Table 3
Calculated scores of attributes in each mobile application service.

	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9
Bump	−1	13	−1	11	1	0	0	52	1
Facebook	5	44	6	15	0	0	1	68	0
Foursquare	4	12	3	5	−1	0	2	25	−1
Fring	2	15	1	12	2	0	0	23	0
Google+	1	5	0	−1	0	0	1	13	0
Skype	0	16	−3	8	0	3	0	38	1
Twitter	10	47	5	15	1	0	3	87	0
WordPress	2	4	−1	9	2	1	0	20	0
Term presence	232	164	199	352	395	184	517	484	309

0 and 1, as shown in Table 4. Also, the *term presence* of attributes in customer reviews was transformed into weights using Eq. (7).

4.5. Evaluation of final level of customer satisfaction

4.5.1. Calculation of ranking of final level of customer satisfaction

VIKOR was applied to evaluate the final level of customer satisfaction. First, according to Table 4, two referential sequences of positive ideal solution f_j^* and negative ideal solution f_j^- were obtained using Eq. (1), as shown in Table 5. Next, the $\frac{w_j\left(f_j^*-f_{ij}\right)}{f_j^*-f_j^-}$ term in Eq. (2) was calculated as shown in Table 6.

Based on the Table 6, Eq. (2) and Eq. (3) results, S_i, R_i and Q_i could be obtained for each service, as shown in Table 7. Here, the

Table 5
Positive ideal solutions f_j^* and negative ideal solutions f_j^-.

	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9
f_j^*	0.144	0.556	0.100	0.100	0.056	0.067	0.067	1.000	0.044
f_j^-	0.022	0.078	0.000	0.000	0.022	0.033	0.033	0.178	0.022

Q_i value of each service was calculated using each v value as $v = 0$, $v = 0.5$, and $v = 1$. According to Table 6 and Table 7, the momentous criterion (Search (C_7) and Touch (C_8)) for improving customer satisfaction with the mobile application services could be found from the R_i values. Besides, it was also found that Google+ and WordPress need to improve service performances for each criterion according to the ranking of the R_i values. Thus, to improve customer satisfaction, each mobile application service must initially pay attention to search (C_7) and touch (C_8). If a decision maker sets the v values as $v = 1$, the Q_i values of each mobile application service, Buimp, Facebook, foursquare, fring, Google+, Skype, Twitter, and WordPress, are 0.662, 0.392, 0.857, 0.705, 1, .728, 0, and 0.767, respectively. Therefore, the ranking of the 8 mobile application services is Twitter > Facebook > Bump > fring > Skype > WordPress > foursquare > Google+. If the decision maker sets the v values as $v = 0.5$, the Q_i values of each mobile application service, Buimp, Facebook, foursquare, fring, Google+, Skype, Twitter, and WordPress, are 0.831, 0.437, 0.761, 0.852, 0.950, 0.864, 0, and 0.883, respectively. In this case, the ranking order of the 8

Table 4
Normalized scores of attributes and weights of attributes.

	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9
Bump	0.022	0.178	0.022	0.156	0.044	0.033	0.033	0.611	0.044
Facebook	0.089	0.522	0.100	0.200	0.033	0.033	0.044	0.789	0.033
Foursquare	0.078	0.167	0.067	0.089	0.022	0.033	0.056	0.311	0.022
Fring	0.056	0.200	0.044	0.167	0.056	0.033	0.033	0.289	0.033
Google+	0.044	0.089	0.033	0.022	0.033	0.033	0.044	0.178	0.033
Skype	0.033	0.211	0.000	0.122	0.033	0.067	0.033	0.456	0.044
Twitter	0.144	0.556	0.089	0.200	0.044	0.033	0.067	1.000	0.033
WordPress	0.056	0.078	0.022	0.133	0.056	0.044	0.033	0.256	0.033
Weight	0.139	0.065	0.124	0.109	0.182	0.171	0.082	0.058	0.070

Table 6
Scores of $w_j(f_j^* - f_{ij})/(f_j^* - f_j^-)$.

	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9
Bump	0.082	0.046	0.055	0.031	0.046	0.065	0.182	0.081	0.000
Facebook	0.037	0.004	0.000	0.000	0.093	0.065	0.122	0.044	0.054
Foursquare	0.045	0.047	0.023	0.078	0.139	0.065	0.061	0.143	0.109
Fring	0.059	0.043	0.039	0.023	0.000	0.065	0.182	0.148	0.054
Google+	0.067	0.056	0.047	0.124	0.093	0.065	0.122	0.171	0.054
Skype	0.074	0.042	0.070	0.054	0.093	0.000	0.182	0.113	0.000
Twitter	0.000	0.000	0.008	0.000	0.046	0.065	0.000	0.000	0.054
WordPress	0.059	0.058	0.055	0.047	0.000	0.043	0.182	0.155	0.054

Table 7
Q values and ranking.

	S_i	R_i	Q_i ($v = 1$)	Rank	Q_i ($v = 0.5$)	Rank	Q_i ($v = 0$)	Rank
Bump	0.587	0.182	0.662	3	0.831	4	1.000	5
Facebook	0.419	0.122	0.392	2	0.437	2	0.482	2
Foursquare	0.710	0.143	0.857	7	0.761	3	0.665	3
Fring	0.614	0.182	0.705	4	0.852	5	1.000	5
Google+	0.799	0.171	1.000	8	0.950	8	0.901	4
Skype	0.629	0.182	0.728	5	0.864	6	1.000	5
Twitter	0.174	0.065	0.000	1	0.000	1	0.000	1
WordPress	0.653	0.182	0.767	6	0.883	7	1.000	5

mobile application services is Twitter > Facebook > foursquare > Bump > fring > Skype > WordPress > Google+. Finally, in the case of $v = 0$, the Q_i values are ranked as follows: Twitter > Facebook > foursquare > Google+ > Bump, fring, Skype, and WordPress, and the Q_i values are 1, 0.482, 0.665, 1, 0.901, 1, 0, and 1, respectively. From this result, the Twitter mobile application service, scoring high customer satisfactions for the various criteria, can be considered as a benchmarking mobile application service for improving customer satisfaction with other mobile application services.

4.5.2. Interpretation of results and sensitivity analysis

Meanwhile, if the v value is 0, the $v\left[\frac{S_i - S^*}{S^- - S^*}\right]$ term in Eq. (4) becomes 0. Thus, the Q_i value is affected, absolutely, only by the R_i value. This fact reveals which criterion should be considered momentously to improve customer satisfaction. From this, one can obtain guidelines for identification of areas for improvement in specific aspects of a service operation. For example, most of the mobile application services require improvement of service performance for search (C_7) and touch (C_8) with reference to Table 6 and the R_i values in Table 7. Additionally, interface (C_2) and message (C_4) do not need much improvement in most of the services. Also, although Twitter does a good job in terms of most of the criteria, post (C_6) should be improved.

Lastly, this study also conducted a sensitivity analysis to rank the influence levels of the v values, as shown in Fig. 5. The rankings of Twitter and Facebook were not at all affected by

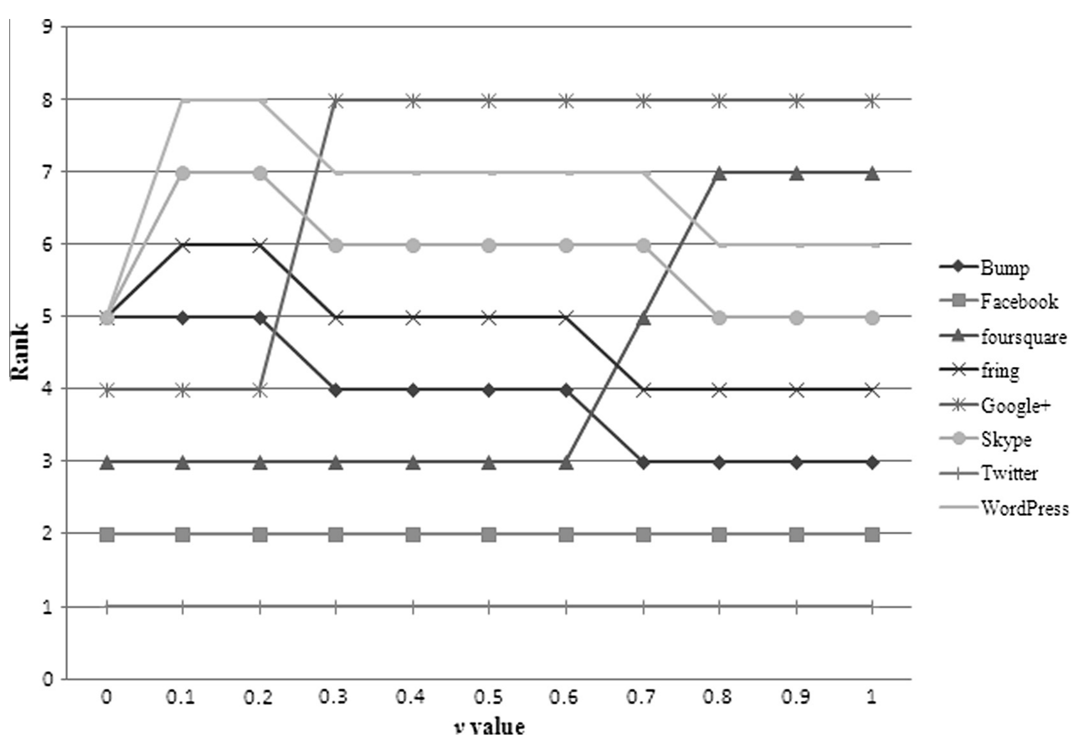

Fig. 5. Sensitivity analysis.

the v value. This means that Twitter and Facebook enjoy high customer satisfaction in terms of both *maximum group utility* and minimum *individual regret*. On the other hand, the rankings of Google+ and foursquare were improved according to decreases of v value. This fact reveals that Google+ and foursquare have higher customer satisfaction when one focuses on minimum *individual regret*. Also, the rankings of Bump, fring, Skype, and WordPress were high when the v value was large, indicating that their rankings were increased when the importance of *maximum group utility* was increased. In other words, they scored high customer satisfaction levels when *maximum group utility* was considered to be important.

5. Conclusions

This study proposed a framework for measuring customer satisfaction in mobile services based on analysis of customer reviews. We presented the technique of sentiment analysis by VIKOR, which entails two main consecutive stages: data collection and preprocessing, and measurement of customer satisfaction. By integrating the strength of sentiment analysis and VIKOR, the proposed approach measures customer satisfaction using actual customer reviews. The results of an empirical case study on mobile application services are presented herein to illustrate the effectiveness of the proposed approach.

The contribution and potential utility of this methodology is twofold. First, traditionally, measurement of the customer satisfaction was conducted using customer surveys. In collecting useful information, much time and effort necessarily is spent. To solve this problem, the proposed approach measures the level of customer satisfaction semi-automatically using sentiment analysis. According to effectiveness, the proposed approach can have similar effects of customer survey which contains customers' view because review documents that are written by customers are utilized in this study. Thus, it can be the basis of further studies that use customer reviews to extract information on customer satisfaction. Second, this study used the MCDM approach, which, significantly, can consider several attributes simultaneously. Notably moreover, VIKOR can provide information on customer satisfaction by considering both *maximum group utility* and *individual regret*. The focus of this study was not limited to the means of measuring customer satisfaction; rather, the proposed approach provides guidelines for identification of areas for improvement in specific aspects of service operation.

Despite all the advantages and possibilities of the proposed approach, it has several limitations that suggest paths for our future research. First, this study used basic techniques of sentiment analysis in formulating the framework for measurement of customer satisfaction presented herein. Therefore, the proposed approach can and should be improved by incorporating more advanced techniques of sentiment analysis. Second, the reliability of the results of the empirical case study presented in these pages was not validated. To that end, a comparative analysis with other MCDM methods should be conducted. Finally, the kind of case study conducted in the present research should be carried out for additional mobile services as well.

Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2011-0030814).

References

Abulaish, M., Jahiruddin & Doja, M. N. (2009). Feature and opinion mining for customer review summarization. *Lectures Notes in Computer Science, 5909*, 219–224.

Beineke, P., Hastie, T., &Vaithyanathan, S. (2004). The sentimental factor: Improving review classification via human provided information. In *Proceedings of Annual Meeting of the Association for Computational Linguistics (ACL)* (pp. 263–270).

Chi, E. H., Prolli, P., & Chen, K. (2001). Using information scent to model user information needs actions on the Web. In *Proceedings of the Human Factors in, Computing Systems* (pp. 490–497).

Choi, J., Seol, H., Lee, S., Cho, H., & Park, Y. (2008). Customer satisfaction factors of mobile commerce in Korea. *Internet Research, 18*(3), 313–335.

Deng, Y., & Chan, F. T. X. (2011). A new fuzzy dempster MCDM method and its application in supplier selection. *Expert Systems with Applications, 38*(8), 9854–9861.

Deng, Z., Lu, Y., Wei, K. K., & Zhang, J. (2010). Understanding customer satisfaction and loyalty: An empirical study of mobile instant messages in China. *International Journal of Information Management, 30*(4), 289–300.

Djoerd, H. (2000). A probabilistic justification for using tf × idf term weighting in information retrieval. *International Journal of Digital Libraries, 3*(2), 131–139.

Haleh, H., & Hamidi, A. (2011). A fuzzy MCDM model for allocating orders to suppliers in a supply chain under uncertainty over a multi-period time horizon. *Expert Systems with Applications, 38*(8), 9076–9083.

Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of ACMSIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 168–177).

Hu, M., & Liu, B. (2004). Mining opinion features in customer reviews. In *Proceedings of the International Conference on Artificial Intelligence, 19*, (pp. 755–760).

Kamps, J., & Marx., M. (2002). Words with attitude. In *Proceedings of 1st International WordNet Conference* pp. (332–341).

Kang, D. K., & Park, Y. T. (2012). Measuring customer satisfaction of service based on an analysis of the user generated contents: sentiment analysis and aggregating function based MCDM approach, In *Proceedings of the 6th IEEE International Conference on Management of Innovation and Technology*.

Kim, M., Park, M., & Jeong, D. (2004). The effects of customer satisfaction and switching barrier on customer loyalty in Korean mobile telecommunication services. *Telecommunications Policy, 28*(2), 145–159.

Kim, H., & Yoon, C. (2004). Determinants of subscriber churn and customer loyalty in the Korean mobile telephony market. *Telecommunications Policy, 28*(9–10), 751–765.

Kuo, Y. F., Wu, C., & Deng, W. (2009). The relationships among service quality, perceived value, customersatisfaction, and post-purchase intention in mobile value-added services. *Computers in Human Behavior, 25*(4), 887–896.

Lee, J., Lee, J., & Feick, L. (2001). The impact of switching costs on the customer satisfaction-loyalty link: Mobile phone service in France. *Journal of Service Marketing, 15*(1), 35–48.

Liou, J. J. H., Tsai, C., Lin, R., & Tzeng, G. (2011). A modified VIKOR multiple-criteria decision method for improving domestic airlines service quality. *Expert Systems with Applications, 17*(2), 57–61.

Liu, B., Hu, M., & Cheng, J. (2005). Opinion observer – Analyzing and comparing opinions on the Web. In *Proceedings of the 14th International Conference on, World Wide Web* (pp. 342–351).

McDonald, R. Hannan, K., & Neylon, T. (2007). Structured models for fine-to-coarse sentiment analysis. In *Proceedings of Annual Meeting of the Association for Computational Linguistics (ACL)* (pp. 432–439).

Mejova, Y. (2009). Sentiment Analysis: An Overview. Comprehensive exam paper, available on http://www.cs.uiowa.edu/~ymejova/publications/CompsYelenaMejova.pdf.

Miller, K. (1998). Modifiers in WordNet. In C. Fellbaum (Ed.), *Word-net: An electronic lexical database*. Cambridge, MA: MIT Press.

Moreo, A., Romero, M., Castro, J. L., & Zurita, J. M. (2012). Lexicon-based comments-oriented news sentiment analyzer system. *Expert Systems with Applications, 39*(10), 9166–9180.

Nasukawa, T., & Yi, J. (2003). Sentiment analysis: capturing favorability using natural language processing. In *Proceedings of the 2nd International Conference on Knowledge Capture* (pp. 494–500).

Opricovic, S. (1998). *Multicriteria optimization of civil engineering systems. Faculty of civil engineering*. Belgrade.

Opricovic, S. (2011). Fuzzy VIKOR with an application to water resources planning. *Expert Systems with Applications, 38*(10), 12983–12990.

Pang, B., & Lee, L. (2004). A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts. In *Proceedings of Annual Meeting of the Association for Computational Linguistics (ACL)* (pp. 271–278).

Pang, B., Lee, L., Vaithyanathan, S. (2002). "Thumbs up? Sentiment classification using machine natural learning techniques. In *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing* (pp. 79–86).

Perry, M., O'Hara, K., Sellen, A., Brown, B., & Harper, R. (2001). Dealing with mobility: Understanding access anytime, anywhere. *ACM Transactions on Computer–Human Interaction, 8*(4), 323–347.

Ponte, J. M., & Croft, W. B. (1998). A language modeling approach to information retrieval. In *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in, Information Retrieval* (pp. 275–281).

Popescu, A. M., Etzioni, O. (2005). Extracting product features and opinions from reviews. In *Proceedings of the 2005 Conference on Empirical Methods in Natural Language Processing* (pp. 339–346).

Poston, R. S., & Speier, C. (2005). Effective use of knowledge management systems: A process model of content ratings and credibility indicators. *MIS Quarterly, 29*(2), 221–244.

Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information Processing and Management, 24*(5), 513–523.

Täckström, O., & McDonald, R. (2011). Semi-supervised latent variable models for sentence-level sentiment analysis. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies* (pp. 569–574).

Turney, P. (2002). Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In *Proceedings of Annual Meeting of the Association for Computational Linguistics (ACL)* (pp. 417–424).

Wang, Y. S., Lin, H. L., & Luarn, P. (2006). Predicting consumer intention to use mobile service. *Information Systems Journal, 16*(2), 157–179.

Wang, Y., & Lo, H. (2002). Service quality, customer satisfaction and behavior intentions: Evidence from China's telecommunication industry. *Info – The Journal of Policy, Regulation and Strategy for Telecommunications, 4*(6), 50–60.

Wilson, T., Wiebe, J., & Hoffmann, P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of Human Language Technology Conference on Empirical Methods in Natural, Language* (pp. 347–354).

Wilson, T., Wiebe, J., & Hoffmann, P. (2009). Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. *Computational Linguistics, 35*(3), 399–433.

Woo, K., & FOCK, H. (1999). Customer satisfaction in the Hong Kong mobile phone industry. *The Service Industries Journal, 19*(3), 162–174.

Zhang, K., Narayanan, R., & Choudhary, A. (2010). Voice of the customers: Mining online customer reviews for product feature-based ranking. *Workshop on Online Social Networks 2010*.

Zhang, W., Xu, H., & Wan, W. (2008). Weakness finder: Find product weakness from Chinese reviews by using aspects based sentiment analysis. *Expert Systems with Applications, 39*(11), 10283–10291.

Zhang, W., Yoshida, T., & Tang, X. (2011). A comparative study of TF∗IDF, LSI and multi-words for text classification. *Expert Systems with Applications, 38*(3), 2758–2765.