Expert Systems with Applications 39 (2012) 11870–11879

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

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

Discovering business intelligence from online product reviews: A rule-induction framework

Wingyan Chung [a,*], Tzu-Liang (Bill) Tseng [b]

[a] UNC Fayetteville State University, School of Business and Economics, 1200 Murchison Road, Fayetteville, NC 28301, USA
[b] Department of Industrial, Manufacturing and Systems Engineering, The University of Texas at El Paso, 500 West University Avenue, El Paso, TX 79968, USA

ARTICLE INFO

Keywords:
E-commerce
Online reviews
Data mining
Text mining
Association rule mining
Rough set theory
Business intelligence
Online reputation

ABSTRACT

Online product reviews are a major source of business intelligence (BI) that helps managers and marketers understand customers' concerns and interests. The large volume of review data makes it difficult to manually analyze customers' concerns. Automated tools have emerged to facilitate this analysis, however most lack the capability of extracting the relationships between the reviews' rich expressions and the customer ratings. Managers and marketers often resort to manually read through voluminous reviews to find the relationships. To address these challenges, we propose the development of a new class of BI systems based on rough set theory, inductive rule learning, and information retrieval methods. We developed a new framework for designing BI systems that extract the relationship between the customer ratings and their reviews. Using reviews of different products from Amazon.com, we conducted both qualitative and quantitative experiments to evaluate the performance of a BI system developed based on the framework. The results indicate that the system achieved high accuracy and coverage related to rule quality, and produced interesting and informative rules with high support and confidence values. The findings have important implications for market sentiment analysis and e-commerce reputation management.

1. Introduction

As e-commerce supports higher interactivity among users with Web 2.0 applications, user-generated content posted on these sites is growing significantly. Users not only consume Web content, but also produce massive data of their participation, often affecting other users' decisions. A study finds that more than three-quarters of the 2078 users reported that online product reviews had a significant influence on their purchase decisions (comScore, 2007). These online product reviews contain descriptions about user preferences, comments, and recommendations that serve as a major source of business intelligence (BI), helping managers and marketers to better understand customers. Management scholar Peter Drucker emphasizes that "what is value to the customer" may be the most important question to answer in order to realize a business's mission and purpose (Drucker, 2003). However, the large volume of online product review data creates significant information overload problems (Bowman, Danzig, Manber, & Schwartz, 1994), making it difficult to discover BI from the reviews and to analyze customer concerns.

Two major pieces of information available in each online review are its textual content and the numerical rating, which respectively indicate the aspects of customer concerns and the customer sentiment. However, neither of these two alone provides the full account of a product's real "value" (Drucker, 2003), which is the true explanation of the customer's satisfaction. An important task of a manager is therefore to correlate between the numerical ratings and the textual content of the reviews in order to understand what the customer values in a product. This task is typically done by manually reading and extracting key phrases or words that indicate customer concerns and by manually relating between the extracted phrases and the numerical ratings. Despite its usefulness, such analysis is time-consuming and does not scale up to the rapidly growing online reviews. Automated tools and techniques have been proposed to analyze online reviews. These works try to study the reviews' impact on sales (Zhu & Zhang, 2010), to recommend products (Aciar, Zhang, Simoff, & Debenham, 2007), to calculate the utility of the reviews (Ding & Liu, 2007), to identify key product features (Zhang, 2008), to detect false reviews (Jindal & Liu, 2007), and to summarize review content (Zhuang, Jing, & Zhu, 2006). However, research that supports the managerial task of correlating between the numerical ratings and textual content of the reviews is not widely found. The problem of how the reviews' textual content contributes to the numerical ratings is thus not widely

* Corresponding author.
E-mail address: wchung@uncfsu.edu (W. Chung).

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2012.02.059

addressed. Understanding this correlation in large amounts of online review data could help e-commerce managers to make effective decisions on brand management, product promotion, and reputation management.

In this paper, we discuss existing works on analyzing online product reviews and critically review these existing approaches. Following a design science paradigm (Hevner, March, Park, & Ram, 2004), we develop a new framework for designing a new class of BI systems that correlate the textual content and the numerical ratings of online product reviews. In contrast to behavioral science, the design science paradigm was chosen because it emphasizes on building and evaluating innovative artifacts that address the analysis needs of e-commerce managers, marketers, and BI practitioners. In the process of building our artifacts, we drew upon the theoretical and computational foundations of data mining (Liu, 2007; Pawlak, 1982) and information retrieval (Salton, 1989; Salton, Wong, & Yang, 1975). Based on the rough set theory and inductive rule mining methods used in the framework, we developed as an instantiation a system for extracting the relationship between hundreds of customer ratings and their corresponding textual reviews posted on Amazon.com's Web site. To demonstrate the applicability of the system, two data mining methods were implemented to extract automatically decision rules to guide the understanding of the relationship. Using quantitative and qualitative experiments, we empirically tested the system that was configured under different methods and settings. The system's enhanced performance was demonstrated over different types of products' online reviews. The results have strong implications for brand management and online market sentiment analysis.

2. BI research and online product review analysis

The term "business intelligence" (BI) is defined as the acquisition, interpretation, collation, analysis, and exploitation of information in business (Chung, Chen, & Nunamaker, 2005; Davies, 2002). BI systems enable organizations to understand their internal and external environments. To support understanding of internal data, one class of BI systems (Carvalho & Ferreira, 2001) manipulates massive operational data to extract essential business information. Some examples of these systems are decision support systems, executive information systems, online-analytical processing (OLAP), data warehouses and data mining systems that are built upon database management systems to reveal hidden trends and patterns. Another class of BI systems tries to systematically collect and analyze information from the external business environment to assist in organizational decision making. They gather information from public sources such as the Internet and provide insights into various knowledge discovery processes. Examples include customer review analysis, Web search log mining, and opinion mining. Technologies to support the second class of BI systems are in general less matured than those for the first class mentioned above.

Research and industry developments in BI have been growing in recent years due to the growing amounts of business data and the widespread use of the Internet as a medium of communication. Many of these works develop Web-based business intelligence systems to assist in data analysis and decision-making (Chung & Chen, 2009; Lawton, 2006). The most recent trends in BI concern about user-generated data analysis. Opinion and sentiment are extracted from large amounts of textual data to facilitate managerial decision-making. There is much room for the design science research community to contribute to this area.

Design science is concerned with the creation and evaluation of new information technology (IT) artifacts with a goal of meeting business needs (Hevner, March, Park, & Ram, 2004). Hevner et al. (2004) provided seven guidelines for design-science research: design as an artifact, problem relevance, design evaluation, research contributions, research rigor, design as a search, and communication of research. While most BI systems are IT artifacts and the BI domain provides sufficient challenges to satisfy the guideline for "problem relevance," there is a need for design science research to address the remaining guidelines mentioned above. The design-science-based information systems research can contribute significantly to BI because the design and evaluation of new IT artifacts within organizational and managerial context can bring new insights about BI technologies, practices, and challenges (Chen, Chiang, & Storey, 2009). Information systems and technologies utilizing data/text/Web mining techniques have been developed to analyze BI from online product reviews.

2.1. Online product review analysis

Data mining and machine learning techniques identify patterns from large amounts of data using statistical and heuristics methods (Mitchell, 1997). These techniques have been applied to a large number of domains, such as business stakeholder classification (Chung, Chen, & Reid, 2009), crime analysis (Chen et al., 2004), and medical data prediction (Brown et al., 2000). Text mining applies data mining techniques to analyzing unstructured, text data (Trybula, 1999). Web mining further uses data and text mining techniques to extract the content, structure, and usage information from Web data (Kosala & Blockeel, 2000). These techniques are applied to a variety of online product review analyses. For example, Zhu and Zhang studied the reviews' impact on sales of online games and found that these reviews are more influential for less popular games and games whose players have greater Internet experience (Zhu & Zhang, 2010). Yan et al. developed a dictionary-based method to represent review textual features and used machine-learning techniques to classify the review sentiment (Dang, Zhang, & Chen, 2010). Zhang used lexical similarity, shallow syntactic features, and lexical subjectivity clues to distinguish useful from useless reviews (Zhang, 2008). To address ambiguity in review text, Ding and Liu used linguistic rules to determine the semantic orientations of words in customer reviews (Ding & Liu, 2007). To support spam detection, reviews were categorized into false opinion (overly positive or negative comments), brand reviews (based only on brand but not product), and non-reviews (advertisements without comment) (Jindal & Liu, 2007). Besides, product recommendation was done through mapping automatically each sentence of a review into a manually-created ontology (Aciar et al., 2007). Different actors were considered in (Zhuang et al., 2006) to summarize movie reviews using WordNet and statistical analysis.

While much of previous research tried to extract sentiment and opinion and to distinguish among different types of product reviews, identifying rules and patterns from online reviews is not widely studied. According to the Merriam-Webster Dictionary, a "pattern" is defined as "a discernible coherent system based on the intended interrelationship of component parts (pattern, 2010)." A rule is defined as "a usually valid generalization" and can be considered a specific type of patterns. Discovering rules from data is a major task in data mining (Liu, 2007), in which a rule is often specified as an association in the form "antecedents ⇒ consequents" such that the left-hand side (e.g., words used in online reviews) of the rule determines the right-hand side (e.g., product rating). These rules are a specific type of patterns that represent associations among any extracted entities. Such rules and patterns often represent valuable knowledge assets in organizations (e.g., tacit knowledge as discussed in p.112 of (Alavi & Leidner, 2001)), providing insights for managerial decision making.

2.2. Inductive rule learning: theories and methods

Inductive rule learning is a major type of machine learning. A rule is represented as a two-part statement with antecedent (LHS) and consequent (RHS), designating a causal relationship between the two parts. As large amounts of data are available from the Internet and other digital storage media, automatic rule-induction methods are used increasingly to extract decision rules from these data. Examples of rule-induction methods include C4.5 (Quinlan, 1993), RIPPER (Cohen, 1995), and CN2 (Clark & Niblett, 1989). Among these methods, association rule mining (Agrawal & Srikant, 1994) and rough-set theory (Pawlak, 1982) are some of the most well-established and widely recognized due to their wide applicability and proven robustness, as described below.

2.2.1. Association rule mining

Association rule mining (ARM) is commonly used in market basket analysis to determine the relationships between items occurring in a large amount of data. The most widely used algorithm for association rule mining is the Apriori algorithm (Agrawal & Srikant, 1994). ARM has been applied to online retailing, product recommendation, and customer relationship management. For instance, Huang et al. found that association rule mining achieved significantly better precision than a high-degree association method (using Hopfield Net) and a direct retrieval method in online product recommendation based on users' transaction data alone (Huang, Chung, & Chen, 2004). Applying association rule mining to customer relationship management in tourism industry, Liao, Chen, & Deng, (2010) extracted association rules indicating the relationship between customer data (demographics, transactions, and preferences) and considerations in new product development. Demiriz compared among association rule mining, dependency network, and another system called e-VZPro using sparse binary data in product recommendation (Demiriz, 2004). Another research applied association rule mining to develop an order batching strategy in supply chain management (Chen, Huang, Chen, & Wu, 2005).

Given its capability in finding associations among a large number of items, ARM can possibly be applied to analyzing the large number of textual features found in online product reviews. However, such application is not widely studied.

2.2.2. Rough-set theory

Rough set theory (RST) (Pawlak, 1982) is a mathematical approach to identifying decision rules from ambiguous and uncertain data, ill-defined problems, indiscernible relations and classifications, and interdependent attributes (Shyng, Wang, Tzeng, & Wu, 2005). The major advantage of RST over standard statistical techniques is the capability to handle qualitative data (Simoudis, Han, & Fayyad, 1996). RST has been applied to many domains, such as fault diagnosis (Dong, Xiao, Liang, & Liu, 2008; Zhang, Shi, & Gao, 2009), interval data clustering (Doumpos, Marinakis, Marinaki, & Zopounidis, 2009; Malcolm & Michael, 2001), supply chain management (Gaudreault, Frayret, & Pesant, 2009; Liang & Huang, 2006), image analysis (Predki, Slowinski, Stefanowski, Susmaga, & Wilk, 2008; Xiao & Zhang, 2008), knowledge acquisition (Jerzy, 1988; Qian, Liang, & Dang, 2008), manufacturing quality control (Huang, Fan, Tseng, Lee, & Huang, 2008; Tseng, Jothishankar, & Wu, 2004), customer relationship management (Tseng & Huang, 2007).

RST has been applied to mining textual data as well. Cheng et al. used RST to achieve a satisfactory reduction of the number of text features (Cheng, Zhang, Wang, & Chen, 2008). Other rule induction approaches in text classification incorporate weight to each feature (e.g., (Wang & Qi, 2009)). Furthermore, many approaches use a hybrid format combining RST and other theories, namely, Fuzzy Set Theory (FST) (Huang, Yang, & Kuo, 2009), Support Vector Machines (SVMs) (Chen & Liu, 2008; Teng, Ren, & Kuriowa, 2007), Naïve Bayes Model (Vidhya & Aghila, 2010), and Chi Square Statistics (Dai, Hu, & Liu, 2008; Li, Li, Liu, & Li, 2004). These hybrid approaches share a common goal of augmenting the existing RST approach to handle different problem domains (Chan, Lin, & Wang, 2009; Nga, Xiu, & Chau, 2009). Despite the widespread applications of RST, it is surprising to find little prior work on applying RST to discovering BI from online product reviews.

3. A rule-induction framework for discovering BI

As the volume of online product reviews grows tremendously in recent years, many efforts have been put in analyzing and understanding these contents. Despite prior works in analyzing these reviews and in mining textual data, identifying rules and patterns from online reviews is not widely studied. These rules and patterns are a basic form of knowledge representation that can enhance managers' and marketers' understanding of their customers. However, existing tools and approaches lack the capability of extracting the relationships between the reviews' rich expressions and the customer ratings. While rule-inductive machine learning has been applied to many domains, surprisingly, no prior work is found on applying association rule mining (ARM) or RST to online product review analysis. Among other techniques that could be applicable, we find ARM and RST to produce the most intuitive results for extracting relationship between customer ratings and online reviews. To address the challenges, we therefore developed a new rule-induction framework and built a BI system based on theories and methods in inductive rule learning, rough set theory, and information retrieval. In the following, we describe the framework and its application to analyzing online product reviews.

3.1. A rule-induction framework

The various components of our framework are shown in Fig. 1. A new IT artifact, the framework was developed to enrich the knowledge base of design science research in information systems (Hevner et al., 2004). The rationale for the design of this framework is to combine the representativeness of key terms in textual reviews and the power of pattern recognition by incorporating techniques from information retrieval (Salton, 1989), data mining (Liu, 2007), and text mining (Chen, 2001; Chung, 2004; Trybula, 1999). The framework was used to develop a BI system to extract key features and to induce rules and patterns from online product reviews. We define a pattern as a significant association among features extracted from the reviews (significance is defined by confidence and other metrics as explained below). A rule is a specific pattern that takes this format: "word 1, word 2, ..., word $n \Rightarrow$ rating" such that the left-hand side (i.e., key words extracted from online reviews) of the rule determines the right-hand side (i.e., product rating). A feature refers to a word appearing in the reviews. *Feature extraction* is the process of automatically identifying features from the reviews. An indexer is needed to document the appearance and frequency of a feature in each review. *Feature filtering* is the process of removing less important features from the set of all features. A heuristic is needed in the process to rank the features such that features with lower ranking scores are removed. As done in previous BI research (Chung, 2008; Chung et al., 2005), we used a list of 462 stop words (such as "a," "and," "the") to remove words that bear little or no semantic meaning. The extracted features are then used as input to RST-based algorithms to induce rules and patterns that represent business intelligence from the online reviews.

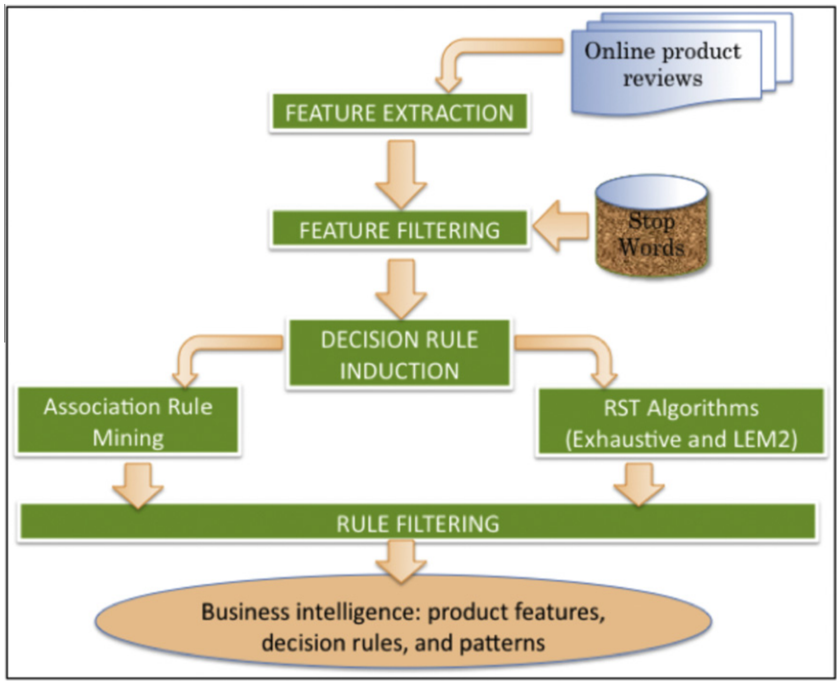

Fig. 1. A rule-induction framework for discovering BI from online product reviews.

3.2. Feature extraction and filtering

Each feature is a textual term appearing in the reviews. Because there are a large number of unique terms in the reviews, we need to select only important terms to serve as features. This selection requires analyzing relationship among the terms and determining the importance of each term. Equation 1 shows the components we considered in calculating the term importance. The two major components in the formula are (1) the normalized term frequency that demonstrates the popularity of a term in a review and (2) the inverse document frequency that demonstrates the specificity of a term in the collection (each document is an online review of a product). These components were based on the theory of term importance from the information retrieval domain (Salton, 1989; Salton, Yang, & Yu, 1975).

Equation 1. Term Importance Formula

$$d_{ij} = \frac{tf_{ij}}{\sum_{k=1}^{n} tf_{ik}^{2}} \times \log\left(\frac{N}{df_j}\right)$$

where, tf_{ij} is the number of occurrence of term j in review i; df_j is the number of reviews containing term j; N is the number of reviews for a product and n is the number of unique textual terms appearing in all the reviews.

The calculation captures the importance of a term in distinguishing a review from the other reviews. The higher the score a term has, the more able the term is in distinguishing a review from another. We then used the term importance scores to rank all extracted features of a product's online reviews and selected the top-ranked features to serve as input of the next steps.

We used as the output classes the numerical ratings (e.g., from 1 to 5) associated with the reviews. Reviewers provide these ratings when they review an item. The textual input features thus serve as antecedents of the decision rules while the output classes serve as consequents. For example, a sample decision rule "useful, excellent ⇒ category = 5" (confidence = 0.9) means that if a review contains the words "useful" and "excellent," then the rating is "5" with a confidence level of 90%.

3.3. Rule-induction methods

Two sets of rule-induction methods – association rule mining and rough-set theory algorithms – were used in our framework to discover decision rules from online product reviews. They were chosen due to their wide applicability in different domains and their robustness in dealing with large number of input features. Each of the rule-induction methods was applied to discovering BI in the form of decision rules from online product reviews.

3.3.1. Association rule mining

The goal of association rule mining is to discover all association rules that have support and confidence higher than the user-specified thresholds. When mining association rules from online reviews, the task is to find association rules that indicate what textual features contribute significantly to certain numerical ratings. Suppose that $F_i = \{f_1, f_2, \ldots, f_m\}$ denote all the textual features (single words) extracted from a set of online reviews, represented by D_i, where each review contains some of the textual features found in F_i. Also let $R = \{1, 2, \ldots, n\}$ be the set of numerical ratings that reviewers provide. Each review can be modeled as a tagged collection of textual features. Therefore, the task of discovering decision rules from the reviews is to identify the association rules in the form: "$X \Rightarrow Y$" where $X \subseteq F_i$ is the set of extracted textual features, $Y \subset R$ is a numerical rating, and $X \cap Y = \varnothing$. To filter the potentially large number of association rules identified from the review, the threshold support and confidence are used. The *support* of the aforementioned rule is the proportion of the reviews containing both feature set X and rating Y. The *confidence* is the proportion of the reviews that are rated as Y if these reviews already contain feature set X. To extract association rules from online reviews, we selected the Apriori Algorithm (Agrawal & Srikant, 1994) in our framework due to the wide applicability and high efficiency of the algorithm. In using the algorithm, we consider the presence or absence (instead of term frequency) of textual features in the reviews because the algorithm handles only binary values of each feature.

Two major steps are involved in the Apriori Algorithm: (1) to find all sets of features that have support exceeding the minimum support threshold (the resulting feature sets are called *frequent feature sets*) and (2) to generate from the frequent feature sets association rules that have confidence exceeding the minimum confidence threshold (Liu, 2007). Because the number of association rules can be very large, filtering is needed in three aspects. First, rules with feature items in the consequents are removed because the consequents should contain only the numerical ratings. Second, rules that contain numerical ratings in the antecedents are removed because only textual features should contribute to the decision of numerical ratings. Third, rules with zero feature values (i.e., absence of a feature) are removed because they are much less meaningful than rules showing presence of a feature.

3.3.2. Rough set theory

Rough set theory (RST) is concerned with measuring the "ambiguity" inherent in the data. A rough set is a collection of objects that cannot be precisely characterized in terms of the values of their sets of attributes, but can be characterized in terms of lower or upper approximations. The upper approximation includes all objects that possibly belong to the concept while the lower approximation contains all objects that definitely belong to the concept. As each object is characterized with attributes, discovering dependencies between attributes and detecting main attributes is of primary importance in RST. Attribute reduction is one unique aspect of the rough set approach. A reduct (Pawlak, 1982) is a minimal sufficient subset of attributes, which provides the same quality of discriminating concepts as the original set of attributes. In a hypothetical example shown in Table 1, there are five reviews each with four input features and an output feature (outcome).

To derive reducts, consider the first feature $F1$. The set of objects corresponding to the feature value $F1 = 0$ is $\{1,2,3,5\}$. This set $\{1,2,3,5\}$ cannot be further classified solely using the relation $F1 = 0$. It is discernible over the constraint $F1 = 0$, which is expressed as $(F1 = 0) = \{1,2,3,5\}$. For the objects in set $\{1,5\}$ the output feature is $O = 2$, for object 3 the output feature is $O = 1$ and for object 2 the output feature is $O = 0$. Therefore additional features are needed to differentiate for $O = 0$, 1, or 2. Applying this concept, the classification power of each feature can be evaluated. For instance, the feature value $F1 = 1$ is specific to $O = 1$. This discernible relation can be extended to multiple features, e.g., $(F1 = 0) \wedge (F2 = 1) = \{1,3\}$ and $(F1 = 0) \vee (F2 = 1) = \{1,2,3,5\}$, where $\wedge$ and $\vee$ refers to "and" and "or" respectively. Most RST-based approaches may generate more than one reduct. Through enumerating all possible reducts, those reducts can be manipulated to produce different decision rules.

Table 1
Example data set.

Review	F_1	F_2	F_3	F_4	O
1	0	1	0	2	2
2	0	0	1	3	0
3	0	1	1	1	1
4	1	2	2	0	1
5	0	0	0	1	2

O: 0 = Not applicable, 1 = Low, 2 = Medium, 3 = High.

Two RST-based algorithms were used in the proposed framework to derive the decision rules. One is called the exhaustive algorithm (Bazan, Nguyen, Nguyen, Synak, & Wróblewski, 2000) while the other is named LEM2 (Learning from Examples Module) algorithm (Grzymała-Busse, 1997). They were chosen due to their effectiveness and high prediction accuracy. The exhaustive algorithm determines the rules (i.e., final reducts) according to the number of reviews supporting the rules with a threshold value and selection from all n-features known rules (n represents number of input features). Since all n-features known rules have been searched, therefore, it is called the exhaustive algorithm. The LEM2 algorithm tries to learn the minimal set of rules to describe the concept. To accomplish this goal, i.e., to learn discriminant description, the LEM2 should be implemented. Moreover, the LEM2 is most frequently used since it often achieves high accuracy.

The rules obtained from the two RST algorithms are then filtered by removing from the data set those rules with fewer than 2 occurrences, a threshold determined empirically.

4. Experimental evaluation

To study the applicability and performance of the proposed framework, we developed a BI system that implements the aforementioned methods and serves as an instantiation of the framework. The goal of the experiment was to study the quality of the rules induced by the system and to compare the performance of different algorithms using a set of evaluation metrics. Our hypotheses in the experiment are that the proposed framework would successfully guide the development of an IT artifact to discover BI in the form of decision rules from online product reviews, that the IT artifact would be applicable to real-world data and would achieve satisfactory performance compared with standards accepted in academic literature.

We conducted an experiment that consists of qualitative and quantitative evaluation of the rules induced by the algorithms. We also built a research test bed consisting of online product reviews posted publicly on the Web site of Amazon.com, one of the largest online retailers in the world. Products on the Amazon Web site have a large number of reviews for a wide range of products. Each review includes a title, a textual description, date, time, author name and location, ratings, and other miscellaneous information. A review is associated with a Likert-scale "star" rating that is one of these values: 1, 2, 3, 4, and 5, where 5-star means "excellent" and 1-star means "poor."

4.1. Experimental test bed and setting

We have tested our system using the online reviews of four products: Acer Aspire One 8.9-inch Mini Laptop (with 686 reviews), Sterling Silver Marcasite & Garnet Glass Heart Pendant 18" (with 78 reviews), Razor Power Wing Caster Scooter (with 78 reviews), and HP 2133-KR922UT 8.9-Inch Mini-Note PC (with 123 reviews). A comprehensive set of online products, these four items represent some of the most popular product categories,

Table 2
Statistics of the product reviews.

No.	Product	Review count	5-star	4-star	3-star	2-star	1-star
1	Acer Aspire One 8.9-inch Mini Laptop	686	425	147	36	21	57
2	Sterling Silver M. & G. Glass Heart Pendant 18"	78	40	24	5	4	5
3	Razor Power Wing Caster Scooter	78	63	10	2	0	3
4	HP 2133-KR922UT 8.9-Inch Mini-Note PC	123	37	39	13	12	22

namely, electronics, action sports, and jewelry. They also have different numbers of reviews that can be used for testing the scalability of the methods. Table 2 shows the basic information about the reviews.

In the feature extraction, ranking, and filtering processes, we selected as features the top 100 terms with the highest term importance scores (calculated using Equation 1) from the reviews for Acer Mini Laptop. For the other three products, we selected the top 40 terms with the highest term importance scores because these three products have relatively fewer reviews. This arrangement helped avoid model over-fitting in the rule-induction process while ensuring representativeness.

We considered as input to the system only feature content (i.e., binary values indicating presence or absence of selected terms) rather than feature values (i.e., term frequencies) based on three reasons. First, each online review is typically a short description consisting of fewer than 100 words. While there are occasional long reviews, most features in a review appear fewer than two times. Therefore, it is more important to capture whether or not a term affects the induced rule than to identify how the frequency of the term affects the induced rule. Second, if both the feature content and value were considered, then there would be too many distinct combinations of content-value of features on the condition part of a rule, creating too many rules that are similar in nature. Third, association rule mining uses only feature content (not value) as data mining input. To make parallel comparison among the algorithms, only feature content should be considered.

The implementation of association rule mining was done by using the Apriori Algorithm (Agrawal & Srikant, 1994) with the following parameters: lower bound of minimum support = 10%, upper bound of minimum support = 100% (reduced by 5% in each iteration), minimum confidence = 50%, and maximum number of rules generated = 200,000. The implementation of the two RST algorithms was done by using the RSES software (available at http://logic.mimuw.edu.pl/~rses/) in a Hewlett Packard XW 4600 Workstation (CPU: Intel Core 2 Dual CPU E6850 3.00 GHz, RAM: 3567 MB).

4.2. Qualitative evaluation: rule induction and validation

We conducted a qualitative evaluation of the rules induced by each algorithm from each product's reviews. For each product, we selected the top five rules from the results and analyzed their content to find any interesting outcomes. If an algorithm induced fewer than five rules from a product's reviews, then all the rules were selected.

In terms of coverage, the rules induced by ARM are supported by a larger number of reviews (as shown in the numbers listed on the LHS of the rules) than the numbers found in rules induced by Exhaustive and LEM2 algorithms. For example, for the product "HP mini-note PC," all the top rules induced by ARM contain keywords that occur in 18 or more reviews while all the top rules induced by the RST algorithms contains keywords occurring in only 5 or fewer reviews. This demonstrates that the patterns found by ARM are more representative than those by the two RST algorithms.

Table 3 shows these rules that are listed in descending order of their confidence values. For example, the rule "battery, love (80) ⇒ 5 (69, 0.86)" (in the "Acer-association rule mining" cell) means that 69 out of 80 reviews that contain the words "battery" and "love" are associated with a rating of "5." The confidence of this rule is 0.86 (=69/80).

We find that all the three algorithms were able to induce decision rules from the data. For example, the rules induced by association rule mining (ARM) related to the Sterling Silver necklace associate the positive keywords "beautiful" and "love" to high ratings. The rules induced by RST exhaustive algorithm identified "long" and "battery" to be keywords leading to high ratings of Acer mini-laptop. The rules induced by RST LEM2 algorithm contain such intuitive adjectives as "easy," "great," and "fun" that indicate clues to high user ratings of the Razor Power Wing Caster Scooter.

In terms of coverage, the rules induced by ARM are supported by a larger number of reviews (as shown in the numbers listed on the LHS of the rules) than the numbers found in rules induced by Exhaustive and LEM2 algorithms. For example, for the product "HP mini-note PC," all the top rules induced by ARM contain keywords that occur in 18 or more reviews while all the top rules induced by the RST algorithms contains keywords occurring in only 5 or fewer reviews. This demonstrates that the patterns found by ARM are more representative than those by the two RST algorithms.

The two RST algorithms have scalability limits. The RST exhaustive algorithm required five to ten seconds to complete the rule-induction computation while LEM2 and ARM typically needed less than a second. The RST LEM2 algorithm failed to induce any rules for Acer mini-laptop due to the relatively larger number of reviews. It demonstrates a general weakness of the algorithm in handling large datasets.

4.3. Quantitative evaluation: rule metrics comparison

The BI system developed based on the proposed framework provides the flexibility of implementing different rule-induction algorithms, helping managers to choose the techniques that work best for their situations. To compare the performances of different algorithms, we conducted a quantitative evaluation of the induced rules by using a set of metrics, namely, the number of rules induced, average support value, Hellinger's divergence, and interest (Berzal et al., 2005). These metrics were chosen to reveal objectively the quality of rules induced, enabling managers to objectively evaluate the techniques to be used in the BI system.

The number of rules (R) induced indicates the count of all rules after the filtering process. The average support value (S) is the average of all the rules' support values. This metric indicates how well the decision rules induced by an algorithm are applicable to understanding the entire set of reviews. The average confidence value (C) is the average of all the rules' confidence values. This metric indicates the extent to which the keywords appearing on the LHS of the rules are associated with specific ratings on the RHS.

Hellinger's divergence (H) measures the amount of information a rule provides (Berzal et al., 2005). For each algorithm, we calculated the average of all the rules' Hellinger's divergence values to find the aggregate amount of information that the set of rules provide. This value can reveal to managers valuable the rules are in providing new information. The formula is given by:

$$H(A \Rightarrow C) = \sqrt{P(A)} \cdot \left[\left(\sqrt{P(A \cap C)} - \sqrt{P(C)} \right)^2 - \left(\sqrt{1 - P(A \cap C)} - \sqrt{1 - P(C)} \right)^2 \right]$$

Interest (I) is calculated by the following formula, which is equivalent to dividing the confidence by the support of a rule (Berzal et al., 2005). In the formula, the more common A and C are, the less interest the rule will have. For each algorithm, we averaged the interest values of all rules to find the extent to which an algorithm produced interesting rules.

$$\text{interest}(A \Rightarrow C) = \frac{P(A \cap C)}{P(A)P(C)} = \frac{P(C|A)}{P(C)}$$

In addition to the above metrics, we also extracted the set of top keywords related to each rating class of a product's reviews.

Table 3
Top rules induced by each algorithm and sample reviews.

Association rule mining	RST – exhaustive	RST – LEM2
Acer Aspire One 8.9-*inch Mini Laptop*		
Battery, love (80) ⇒ 5 (69, 0.86)	Great, work, computer, product (2) ⇒ 5 (2, 1)	N. A.
My, love (124) ⇒ 5 (102, 0.82)	Great, want, computer (7) ⇒ 5 (5, 0.71)	
Great, love (90) ⇒ 5 (74, 0.82)	Battery, computer, works, long (3) ⇒ 4 (2, 0.66)	
Love (145) ⇒ 5 (119, 0.82)	ACER (179) ⇒ 5 (109, 0.609)	
Laptop, hours (92) ⇒ 5 (74, 0.8)	Screen, product (24) ⇒ 5 (13, 0.54)	
Sample reviews	*Sample reviews*	*N.A.*
The 6 cell **battery** life is awesome ... You'll **love** the sleek shiny white cover.	**Great** little **computer**, very compact and easy to use **product**.	
Ultra-portable device, long last **battery**, ... **love** the quality of camera.	It is **great** for people on the go, business... it is **computer** I choose to ...	
But it works great ... **love** the **battery** that lasts much longer ...	I think my **Acer** is great. I take it everywhere with me.	
Sterling Silver M. & G. Glass Heart Pendant 18"		
Love (12) ⇒ 5 (9, 0.75)	Stone (2) ⇒ 2 (2, 1)	Darker, picture, necklace (3) ⇒ 5 (3, 1)
Beautiful, necklace (14) ⇒ 5 (10, 0.71)	Girlfriend, necklace (2) ⇒ 5 (2, 1)	My, beautiful (3) ⇒ 5 (3, 1)
Beautiful (27) ⇒ 5 (18, 0.67)	Color, red, heart (2) ⇒ 5 (2, 1)	My, bought, picture (3) ⇒ 4 (3, 1)
Chain, beautiful (12) ⇒ 5 (8, 0.67)	Piece (11) ⇒ 5 (5, 0.45)	Chain, darker, picture (3) ⇒ 4 (3, 1)
My, beautiful (17) ⇒ 5 (11, 0.65)	Piece (11) ⇒ 4 (4, 0.36)	Wife, my, pendant, chain (2) ⇒ 5 (2, 1)
Sample reviews	*Sample reviews*	*Sample reviews*
I have this **necklace** and I **love** it. ... Simply **beautiful**!	Gorgeous **necklace** ... Just what any **girlfriend** would like ...	Not quite as **picture**, but I love it ... The heart is definitely **darker** than **picture**.
It is as **beautiful** as it looks. ... The twisted **chain** really sets the pendent off.	I get compliments on it when I wear it, but the **stone** is darker ...	I think its **beautiful** but for the money ...
I bought this for **my** girlfriend ... It's very **beautiful**, it looks very classy.	The design is very intricate, and really makes the **piece** standout ...	I bought this for **my wife** and she has had it for two months.
Razor Power Wing Caster Scooter		
My, easy (17) ⇒ 5 (17, 1)	My, year, scooter (2) ⇒ 5 (2, 1)	Year, great (9) ⇒ 5 (9, 1)
My, scooter, great (15) ⇒ 5 (15, 1)	Great (20) ⇒ 5 (19, 0.95)	Powerwing (7) ⇒ 5 (7, 1)
Fun, great (13) ⇒ 5 (13, 1)	Fun, kids (7) ⇒ 5 (6, 0.857)	Fun, my, easy (5) ⇒ 5 (5, 1)
Fun, easy (12) ⇒ 5 (12, 1)	My (31) ⇒ 5 (26, 0.838)	My, scooter (5) ⇒ 5 (5, 1)
Good, my ⇒ 5 (12, 1)	Fun, my (12) ⇒ 5 (10, 0.833)	My, product (5) ⇒ 5 (5, 1)
Sample reviews	*Sample reviews*	*Sample reviews*
I bought this for **my** 4 yr old son ... This is a awesome toy or **scooter** ... It really is a **great** toy and well worth the price ...	We bought the **scooter** for **my** 5 **year** old daughter. The whole family loves it.	The **Powerwing** is fun for younger kids...
Very **easy** to put together ... Great price and Great **fun** !!	Way too much **fun**! Sturdy!	**Great** product! Our 7 **year**s old Daughter save her pocket money...
It's pretty **easy** to manouver and a heck of a lot of **fun**.	I bought this for **my** 6 yr old daughter...she absolutely loves it! This is a **great** and durable scooter. It adds a spin to the traditional scooter.	It's **easy** for **my** kids to operate and control. We really have a lot of **fun**!
HP 2133-KR922UT 8.9-*Inch Mini-Note PC*		
Machine, me (18) ⇒ 4 (12, 0.67)	note, great (3) ⇒ 5 (2, 0.67)	slow, linux (3) ⇒ 1 (3, 1)
My, vista, screen (18) ⇒ 4 (12, 0.67)	my, XP(5) ⇒ 4 (3, 0.6)	system, keyboard, screen, xp (3) ⇒ 4 (3, 1)
Vista, screen (22) ⇒ 4 (14, 0.64)	my, XP(5) ⇒ 5 (2, 0.4)	Vista, nice (3) ⇒ 4 (3, 1)
Battery, screen (20) ⇒ 4 (12, 0.6)		battery, Vista (3) ⇒ 4 (3, 1)
Vista, machine (21) ⇒ 4 (12, 0.57)		my, xp (5) ⇒ 4 (3, 0.6)
Sample reviews	*Sample reviews*	*Sample reviews*
I like the **machine**, use it every day, but would like more **battery** life and a 10" **screen** like the ASUS which costs much less.	I got the mini **note** hoping it is **great**!	After suffering with **Vista** for a month, I installed Windows XP and was very **nice**.
The keyboard is like fullsize to **me** ... I highly recommend this **machine**.	According to **my** experience, Window **XP** is more stable!	The problem is, I am **slow** with **Linux** and I am totally lost.
The **screen** ... will tend to be a bit too much of a mirror ... The pre-installed **vista** is TOOOOO SLOOOOOWWWWW.	**My** only issue involved the choice I made in O.S. With **XP** pro, including Sp 3 I rate it 5 stars. Amazon is my choice in Re-seller.	... I am using **Vista** now and the **battery** power on the computer does not last as long.

These keywords were identified by first categorizing the rules (identified by each algorithm for each product's reviews) based on the rating, and then were extracted from the left-hand side of the rules. Marketers and managers would be able to use these keywords to explain customers' concerns and sentiment. This metric would enable decision makers to understand at an aggregate level which words are leading to which ratings. To further discern the importance of these keywords, we developed a word-rating score (R) to rank the keywords based on this formula:

$$R_i = \sum_j \frac{\text{confidence}(Rule_{ij})}{|LHS(Rule_{ij})|}$$

In the formula, $Rule_{ij}$ refers to the rule j in which the word i appears on its left-hand side (LHS). The numerator is the confidence value of the rule while the denominator is the number of words found on the LHS of the rule. Intuitively, a word will have a high word-rating score if it dominates the LHS of rules having high confidence values. Table 4 shows the results of the quantitative evaluation of the algorithms.

Table 4
Quantitative evaluation results of the algorithms. A bold value in a cell means it is the highest value of the same metric (R, S, C, I, H) across the three techniques (ARM, RST-E, RST-L).

Algorithm	Association rule mining					RST – exhaustive					RST – LEM2				
Product	R	S	C	I	H	R	S	C	I	H	R	S	C	I	H
Acer	**192**	**0.138**	**0.700**	8.42	0.004	26	0.068	0.481	**81.4**	**0.023**	187	N/A	N/A	N/A	N/A
Sterling	16	**0.115**	0.597	9.74	0.011	5	0.072	0.764	24.6	**0.032**	**23**	0.028	**1.000**	**36.7**	0.027
Scooter	**168**	**0.154**	0.909	23.63	0.018	34	0.117	0.728	13.8	0.027	21	0.046	**1.000**	**26.4**	**0.037**
HP	16	**0.113**	0.561	5.456	0.0005	3	0.035	0.556	17.3	**0.057**	**34**	0.018	**0.978**	**56.1**	0.018

Table 5
Top keywords (and word-rating score) categorized by rating class.

Rating	Association rule mining	RST - exhaustive	RST – LEM2
Acer Aspire One 8.9-*inch Mini Laptop*			
5	My (19.6), battery (14.5), great (10.4), acer (8.0), keyboard (7.9), laptop (7.3), aspire (5.7), small (5.6), computer (4.5), drive (4.4)	Acer (15.6), notebook (15.3), battery (12.79), great (10.9), computer (8.4), work (5.3)	N/A
4	None	Battery (9.5), work (6.7), long (4.7)	N/A
Sterling Silver M. & G. Glass Heart Pendant 18"			
5	Beautiful (1.69), my (1.68), chain (1.44), necklace (1.15), pendant (1.09), love (0.75)	Necklace (1.73), girlfriend (1.58), color (1.1), piece (0.82)	My (2.1), beautiful (1.87), necklace (1.76), picture (1.66), chain (1.63), pendant (1.56) , wife (1.3), darker (0.92)
4	Bought (0.845), darker (0.63), my (0.285)	Piece (1.12)	My (1.72), picture (1.32), chain (0.91), bought (0.72)
2	None	Stone (1.32)	None
Razor Power Wing Caster Scooter			
5	My (26.6), great (14.3), scooter (13.7), year (11.8), fun (10.6), ride (9.7), kids (8.9), son (8.6), powerwing (7.2), bought (6.0)	My (23.4), great (16.12), year (12.22), scooter (11.2), fun (8.92), kids (5.61)	My (24.5), year (14.15), great (11.34), scooter (10.56), fun (9.6), easy (7.34), powerwing (6.4), product (4.32)
HP 2133-KR922UT 8.9-*Inch Mini-Note PC*			
5	Great (0.76), xp (0.26)	Great (1.83), note (0.96), xp (0.82)	None
4	Vista (2.4), my (0.9), machine (0.9), screen (0.8), battery (0.8), me (0.6), nice (0.5), keyboard (0.29), laptop (0.28), windows (0.25), hp (0.17)	My (0.87), xp (0.74)	Vista (3.2), my (2.4), screen (1.3), nice (1.24), battery (1,1), xp (0.92), system(0.7) , keyboard (0.3)
1	None	None	Linux (0.88), slow (0.56)

Only the rating with at least one keyword are displayed. For example, ratings 1 and 3 are not displayed for the product "Sterling Silver M. & G. Glass Heart Pendant 18" because there are not enough product reviews to generate sufficient decision rules and their associated keywords.

The quantitative metrics show that association rule mining (ARM) obtained the highest values in terms of support across all four products' reviews. For the product "Acer mini-laptop" that has the largest number of reviews, ARM obtained the highest values in the number of rules, support and confidence. The number of rules induced from Razor Power Scooter's reviews by ARM is significantly larger than those of RST exhaustive and LEM2 algorithms. These results indicate that the rules induced by ARM are able to represent the most reviews.

Among three of the four products' reviews (Acer, Sterling, and HP), the RST exhaustive algorithm induced rules that have the highest Hellinger's divergence and interest values, while the RST LEM2 algorithm has the best value of the same metric for Razor Power Scooter. These results indicate that the rules induced by RST algorithms are the most informative and interesting. However, LEM2 was not able to compute all the four metrics for Acer's reviews due to the large volume.

In addition to using these metrics, we listed the key terms that have high impact on the rule induction process. Table 5 shows the keywords and word-rating scores categorized by the product rating class (from 1-star to 5-star). Because most of the reviews are classified into one or two rating classes, most keywords are clustered in these rating classes while no keyword was identified for less-frequently-used classes.

5. Discussion

5.1. Managerial application of the BI system

To explicate the managerial application of the BI system, we describe below a hypothetical case that demonstrates the use of the induced rules by the manager of a major online retailer. The manager was used to conducting manual analysis of online product reviews. In such analysis, he reads a large number of these reviews and documented in a spreadsheet file key phrases, words, and rating from each review. While the spreadsheet program allows him to calculate an average of the numerical ratings and to search among the key words, the manager is unable to identify significant patterns in the form of rules that signal correlation between extracted key terms and customer sentiment. The numerical ratings also fail to convey customers' concerns because these numbers over-simplify the more complicated concerns expressed in the reviews. They fail to explain what attributed to the customer dissatisfaction. For example, customer dissatisfaction on certain product features cannot be found from the customer rating. However, using only the textual reviews may not indicate the level of satisfaction or dissatisfaction of customers.

Recently, the manager was introduced to the new BI system that can identify automatically decision rules from online reviews. Each rule is in the form of "antecedents ⇒ consequent (strength)" where the antecedents consist of words extracted from the reviews, the consequent is a numerical rating, and the strength is a number between 0 and 1 where a number closer to 1 means that the rule has a high confidence. In analyzing the reviews for the product "Razor Power Wing Caster Scooter," the manager found that, among the rules that indicate a 5-star rating, some of the key words found in the reviews are "easy," "my," "kids," and "Christmas." These words signal several factors may lead to the high rating: (1) ease of operations of the scooter by kids, (2) sales support during the Christmas season, and (3) promotion strategy targeting parents who have kids. While these factors may be obtained from manual analysis as well, the BI system assist the manager in finding these factors so that he can spend his valuable time in addressing these factors.

Using the system, the manager is able to explain part of the reasons behind the customers' ratings. By correlating between the numerical ratings and the review text, customers' sentiment of the items being reviewed can be analyzed and explained more clearly. The benefits to management include higher efficiency, saving in time, and new insights that may not be available from manual analysis. The system also provides clues in the form of keywords. Independent from numerical ratings, these keywords help to identify customers' concerns and salient topics.

5.2. Discussions and implications

The experimental results provide important implications for researchers, managers, marketers, and system developers. While online product reviews have been widely used in e-commerce and other Web sites, to our knowledge this research is the first attempt to apply rule-induction techniques to discover business intelligence from online product reviews. The integration of information retrieval, data mining, and text mining into a comprehensive framework provides new insights to the related research fields. The intuitive decision rules generated by the BI system demonstrate the novelty and applicability.

As e-commerce continues to grow rapidly, managers and marketers should find the results helpful in their analysis and understanding of market sentiment. For example, the keywords identified in the decision rules provide important clues to correlating between customers' concerns and product ratings. Managers are then able to use these keywords and rules strategically to enhance their products and to manage customer relationship. Marketers can promote their products on the Internet by strategically place the keywords in Web pages so as to increase their sites' ranking in search engines. Search engine optimization is thus supported by using the keywords and rules.

Specifically, managers and marketers should find association rule mining (ARM) algorithm inducing more interesting rules, generating intuitive ideas, while RST algorithms producing more informative rules for deeper analysis. For system developers, ARM algorithm should deliver higher scalability and efficiency than the two RST algorithms. Thus ARM should be used when a large amount of product reviews need to be processed or when time is a critical issue in computation.

6. Conclusions and future directions

A major source of business intelligence, online product reviews impact directly on company reputation and brand perception that can be translated to substantial changes in market share and profitability. Automated tools have emerged to support the analysis of these reviews, however most lack the capability of extracting the relationships between the reviews' rich expressions and the customer ratings. Based on the theories and methods in inductive rule learning, rough set theory (RST), and information retrieval, we developed and validated a new framework for discovering business intelligence from online product reviews. A BI system was developed as an instantiation of the framework to induce automatically decision rules that relate the keywords occurring in online reviews and the customer ratings. Using the reviews of four products sold on Amazon.com, we experimented with the association rule mining (ARM) method, RST exhaustive algorithm, and RST LEM2 algorithm to study how they contribute differently to the quality of decision rules. The experimental results suggest that ARM algorithm achieved the best scalability and efficiency, and induced rules that have the highest level of support and highest confidence for the product with the largest number of reviews, while RST algorithms produced rules that are the most informative, most interesting, and have the highest confidence values. These IT artifacts provide new tools to managers and marketers to analyze their rapidly-growing online product reviews. The results have important implications for market sentiment analysis, online reputation management, and search engine optimization. Our evaluation methodology, which includes a new metric called "word-rating score," provides new guidelines for future research to evaluate BI analytics.

Future work can consider expanding the review datasets to cover different types of product, testing other methods for BI discovery, and studying reviews with varied distributions of ratings. Considering the rapid growth of e-commerce and the widespread use of online product reviews, companies that lack the capability of efficient and effective analysis of these reviews could lose significant competitive advantage.

Acknowledgments

This research is part of the outcome of the project titled "Establishment of the Knowledge Systems Laboratory" (http://ksl.uncfsu.edu/) funded by UNC Fayetteville State University Title III/CCRA Grant. We thank all the project participants for their help.

References

Aciar, S., Zhang, D., Simoff, S., & Debenham, J. (2007). Informed recommender: Basing recommendations on consumer product reviews. *IEEE Intelligent Systems, 22*, 39–47.

Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules. In J. B. Bocca, M. Jarke, & C. Zaniolo (Eds.), *Proceedings of the 20th International Conference on Very Large Data Bases* (pp. 487–499). Santiago, Chile: Morgan Kaufmann.

Alavi, M., & Leidner, D. E. (2001). Review: Knowledge management and knowledge management systems: Conceptual foundations and research issues. *MIS Quarterly, 25*, 107–136.

Bazan, J., Nguyen, H. S., Nguyen, S. H., Synak, P., & Wróblewski, J. (2000). Rough set algorithms in classification problem. In L. Polkowski, S. Tsumoto, & T. Lin (Eds.), *Rough Set Methods and Applications* (pp. 49–88). Heidelberg New York: Physica-Verlag.

Berzal, F., Cubero, J.-C., Marin, N., Sanchez, D., Serrano, J.-M., & Vila, A. (2005). Association rule evaluation for classification purposes. In *Actas del III Taller Nacional de Mineria de Datos y Aprendizaje* (pp. 135–144). Granada, Spain: Thomson.

Bowman, C. M., Danzig, P. B., Manber, U., & Schwartz, F. (1994). Scalable internet resource discovery: Research problems and approaches. *Communications of the ACM, 37*, 98–107.

Brown, M. P., Grundy, W. N., Lin, D., Cristianini, N., Sugnet, C. W., Furey, T. S., Ares, M., & Haussler, D. (2000). Knowledge-based analysis of microarray gene expression data by using support vector machines. In *Proceedings of the national academy of science* (Vol. 97, pp. 262–267).

Carvalho, R., & Ferreira, M. (2001). Using information technology to support knowledge conversion processes. *Information Research, 7*, 23.

Chan, C. W., Lin, C. T., & Wang, L. Q. (2009). Mining the text information to optimizing the customer relationship management. *Expert Systems with Applications, 36*, 1433–1443.

Chen, H. (2001). *Knowledge management systems: A text mining perspective*. Tucson, AZ: The University of Arizona.

Chen, H., Chiang, R. H. L., & Storey, V. C. (2009). MIS quarterly special issue on business intelligence research. In *MIS central* (Vol. 2009), http://www.misq.org/.

Chen, H., Chung, W., Xu, J. J., Wang, G., Chau, M., & Qin, Y. (2004). Crime data mining: A general framework and some examples. *IEEE Computer, 37*, 50–56.

Chen, M.-C., Huang, C.-L., Chen, K.-Y., & Wu, H.-P. (2005). Aggregation of orders in distribution centers using data mining. *Expert Systems with Applications, 28*, 453–460.

Chen, P., & Liu, S. (2008). Rough set-based SVM classifier for tex categorization. In *The fourth international conference of natural computation, 2008* (Vol. 2, pp. 153–157). IEEE Computer Society.

Cheng, Y., Zhang, R., Wang, X., & Chen, Q. (2008). Text Feature Extraction Based on Rough Set. In The *fifth international conference of fuzzy systems and knowledge discovery 2008* (Vol. 2, pp. 301-314). Shandong, China: IEEE Computer Society.

Chung, W. (2004). An automatic text mining framework for knowledge discovery on the web. Unpublished Ph.D. dissertation, The University of Arizona, Tucson, AZ.

Chung, W. (2008). Visualizing E-business stakeholders on the web: A methodology and experimental results. *International Journal of Electronic Business, 6*, 25–46.

Chung, W., & Chen, H. (2009). Web-based business intelligence systems: A review and case studies. In G. Adomavicius & A. Gupta (Eds.). *Handbooks in Information Systems: Business Computing* (Vol. 3, pp. 373–396). Emerald Group Publishing Limited.

Chung, W., Chen, H., & Nunamaker, J. F. (2005). A visual framework for knowledge discovery on the web: An empirical study on business intelligence exploration. *Journal of Management Information Systems, 21*, 57–84.

Chung, W., Chen, H., & Reid, E. (2009). Business stakeholder analyzer: An experiment of classifying stakeholders on the web. *Journal of the American Society for Information Science and Technology, 60*, 59–74.

Clark, P., & Niblett, T. (1989). The CN2 induction algorithm. *Machine Learning, 3*, 261–283.

Cohen, W. (1995). Fast effective rule induction. In *Proceedings of the twelfth international conference on machine learning* (pp. 115–123). Lake Tahoe, CA: Morgan Kaufmann.

comScore, & The Kelsey Group. (2007). Online consumer-generated reviews have significant impact on offline purchase behavior. In http://www.comscore.com/Press_Events/Press_Releases/2007/11/Online_Consumer_Reviews_Impact_Offline_Purchasing_Behavior.

Dai, L., Hu, J., & Liu, W. (2008). Using modified CHI square and rough set for text categorization with many redundant features. In *2008 international symposium of the computational intelligence and design* (Vol. 1, pp. 182–185). Wuhan, China: IEEE Computer Society.

Dang, Y., Zhang, Y., & Chen, H. (2010). A lexicon-enhanced method for sentiment classification: An experiment on online product reviews. *IEEE Intelligent Systems, 25*, 46–53.

Davies, P. H. J. (2002). Intelligence, information technology, and information warfare. In M. E. Williams (Ed.). *Annual Review of Information Science and Technology* (Vol. 36, pp. 313–352). Medford, NJ: Information Today, Inc.

Demiriz, A. (2004). Enhancing product recommender systems on sparse binary data. *Data Mining and Knowledge Discovery, 9*, 147–170.

Ding, X., & Liu, B. (2007). The utility of linguistic rules in opinion mining. In *Proceedings of the 30th annual international ACM SIGIR conference on research and development in information retrieval* (pp. 811–812). Amsterdam, The Netherlands: ACM New York, NY, USA.

Dong, L., Xiao, D., Liang, Y., & Liu, Y. (2008). Rough set and fuzzy wavelet neural network integrated with least square weighted fusion algorithm based fault diagnosis research for power transformers. *Electric Power Systems Research, 78*, 129–136.

Doumpos, M., Marinakis, Y., Marinaki, M., & Zopounidis, C. (2009). An evolutionary approach to construction of outranking models for multicriteria classification: The case of the ELECTRE TRI method. *European Journal of Operational Research, 199*, 496–505.

Drucker, P. (2003). *The daily drucker*. New York, NY: Harper Collins.

Gaudreault, J., Frayret, J. M., & Pesant, G. (2009). Distributed search for supply chain coordination. *Computers in Industry, 60*, 441–451.

Grzymała-Busse, J. (1997). A new version of the rule induction system LERS. *Fundamenta Informaticae, 31*, 27–39.

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *Management Information Systems Quarterly, 28*, 75–105.

Huang , C. C., Fan, Y. N., Tseng , T. L., Lee, C. H., & Huang, H. F. (2008). A hybrid data mining approach to quality assurance of manufacturing process. In *IEEE international conference on fuzzy systems* (pp. 818–825). Hong Kong, China: Institute of Electrical and Electronics Engineers.

Huang, H. H., Yang, H. C., & Kuo, Y. H. (2009). A fuzzy-rough hybrid approach to multi-document extractive summarization. In *Ninth international conference of hybrid intelligent systems, 2009* (pp. 168–173). Shenyang, China: IEEE Computer Society.

Huang, Z., Chung, W., & Chen, H. (2004). A graph model for e-commerce recommender systems. *Journal of the American Society for Information Science and Technology, 55*, 259–274.

Jerzy, W. G. (1988). Knowledge acquisition under uncertainty – a rough set approach. *Journal of Intelligent and Robotic Systems, 1*, 3–16.

Jindal, N., & Liu, B. (2007). Analyzing and detecting review spam. In *Proceedings of the 7th IEEE international conference on data mining* (pp. 547–552). Omaha, NE: IEEE Press.

Kosala, R., & Blockeel, H. (2000). Web mining research: A survey. *ACM SIGKDD Explorations, 2*, 1–15.

Lawton, G. (2006). Making business intelligence more useful. *IEEE Computer, 39*, 14–16.

Li, Q., Li, J. H., Liu, G. S., & Li, S. H. (2004). A rough set-based hybrid feature selection method for topic-specific text filtering. In *2004 international conference of the machine learning and cybernetics* (Vol. 3, pp. 1464–1468). Shanghai Jiao Tong Univ., Shanghai, China: IEEE Computer Society.

Liang, W. Y., & Huang, C. C. (2006). Agent-based demand forecast in multi-echelon supply chain. *Decision Support Systems, 42*, 390–407.

Liao, S.-h., Chen, Y.-J., & Deng, M.-y. (2010). Mining customer knowledge for tourism new product development and customer relationship management. *Expert Systems with Applications, 37*, 4212–4223.

Liu, B. (2007). *Web data mining: Exploring hyperlinks, contents, and usage data*. Berlin Heidelberg: Springer-Verlag.

Malcolm, J. B., & Michael, J. P. (2001). Variable precision rough set theory and data discretisation: An application to corporate failure prediction. *Omega, 29*, 561–576.

Mitchell, T. M. (1997). *Machine learning*. New York: McGraw-Hill.

Nga, E. W. T., Xiu, I., & Chau, D. C. K. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. *Expert Systems with Applications, 36*, 2592–2602.

pattern. (2010) Merriam-Webster online dictionary: http://www.merriam-webster.com/dictionary/pattern, retrieved on August 14.

Pawlak, Z. (1982). Rough sets. *International Journal of Computer and Information Sciences, 11*, 341–356.

Predki, B., Slowinski, R., Stefanowski, J., Susmaga, R., & Wilk, S. (2008). ROSE – software implementation of the rough set theory. Lecture Notes in Computer Science, 1424.

Qian, Y., Liang, J., & Dang, C. (2008). Converse approximation and rule extraction from decision tables in rough set theory. *Computers & Mathematics with Applications, 55*, 1754–1765.

Quinlan, J. R. (1993). *C4.5: programs for machine learning*. Los Altos, CA: Morgan Kaufmann.

Salton, G. (1989). *Automatic text processing: The transformation, analysis, and retrieval of information by computer*. Reading, MA: Addison-Wesley.

Salton, G., Wong, A., & Yang, C. S. (1975a). A vector space model for automatic indexing. *Communications of the ACM, 18*, 613–620.

Salton, G., Yang, C. S., & Yu, C. T. (1975b). A theory of term importance in automatic text analysis. *Journal of the American Society for Information Sciences, 26*, 33–44.

Shyng, J. Y., Wang, F. K., Tzeng, G. H., & Wu, K. S. (2005). Rough set theory in analyzing the attributes of combination values for the insurance market. *Expert Systems with Applications, 32*, 56–64.

Simoudis, E., Han, J., & Fayyad, U. (1996). *Proceedings of the second international conference on knowledge discovery and data mining*. Menlo Park, CA: AAAI Press.

Teng, Z., Ren, F., & Kuriowa, S. (2007). Emotion recognition from text based on the rough set theory and the support vector machines. In *2007 international conference of the natural language processing and knowledge engineering* (pp. 36–41). Beijing, China: IEEE Computer Society.

Trybula, W. J. (1999). Text mining. In M. E. Williams (Ed.). *Annual review of information science and technology* (Vol. 34, pp. 385–419). Medford, NJ: Information Today, Inc.

Tseng, T. L., & Huang, C. C. (2007). Rough set-based approach to feature selection in customer relationship management. *Omega, 35*, 365–383.

Tseng, T. L., Jothishankar, M. C., & Wu, T. T. (2004). Quality control problem in printed circuit board manufacturing – An extended rough set theory approach. *Journal of Manufacturing Systems, 23*, 56–72.

Vidhya, K. A., & Aghila, G. (2010). Hybrid text mining model for document classification. In *The 2nd international conference of the computer and automation engineering (ICCAE)* (Vol. 1, pp. 210–214). Singapore: IEEE Computer Society.

Wang, C. L., & Qi, Y. M. (2009). Variable precision rough set weight calculation based on web text classification. In *The 5th international conference of the wireless communications, networking and mobile computing, 2009* (pp. 1–4.). Beijing, China: IEEE Computer Society.

Xiao, H., & Zhang, X. (2008). Comparison studies on classification for remote sensing image based on data mining method. *WSES Transactions on Computers, 7*, 552–558.

Zhang, Z. (2008). Weighing stars: Aggregating online product reviews for intelligent e-commerce applications. *IEEE Intelligent Systems, 23*, 42–49.

Zhang, Z., Shi, Y., & Gao, G. (2009). A rough set-based multiple criteria linear programming approach for the medical diagnosis and prognosis. *Expert Systems with Applications, 36*, 8932–8937.

Zhu, F., & Zhang, X. M. (2010). Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. *Journal of Marketing, 74*, 23.

Zhuang, L., Jing, F., & Zhu, X.-Y. (2006). Movie review mining and summarization. In *Proceedings of the 15th ACM international conference on Information and knowledge management* (pp. 43–50). Arlington, Virginia, USA: ACM New York, NY, USA.