

## Association for Information Systems AIS Electronic Library (AISeL)

---

ICIS 2010 Proceedings

International Conference on Information Systems  
(ICIS)

---

2010

# Extracting Business Intelligence from Online Product Reviews: An Experiment of Automatic Rule-Induction

Wingyan Chung

*Santa Clara University*, [wchung@scu.edu](mailto:wchung@scu.edu)

Tzu-Liang (Bill) Tseng

*The University of Texas at El Paso*, [btseng@utep.edu](mailto:btseng@utep.edu)

Follow this and additional works at: [http://aisel.aisnet.org/icis2010\\_submissions](http://aisel.aisnet.org/icis2010_submissions)

---

### Recommended Citation

Chung, Wingyan and Tseng, Tzu-Liang (Bill), "Extracting Business Intelligence from Online Product Reviews: An Experiment of Automatic Rule-Induction" (2010). *ICIS 2010 Proceedings*. 233.

[http://aisel.aisnet.org/icis2010\\_submissions/233](http://aisel.aisnet.org/icis2010_submissions/233)

This material is brought to you by the International Conference on Information Systems (ICIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in ICIS 2010 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact [elibrary@aisnet.org](mailto:elibrary@aisnet.org).

# Extracting Business Intelligence from Online Product Reviews: An Experiment of Automatic Rule-Induction

*Research-in-Progress*

**Wingyan Chung**

Department of Operations and  
Management Information Systems,  
Santa Clara University,  
Santa Clara, California, USA  
wchung@scu.edu

**Tzu-Liang (Bill) Tseng**

Department of Industrial, Manufacturing,  
and Systems Engineering,  
The University of Texas at El Paso,  
El Paso, Texas, USA  
btseng@utep.edu

## **Abstract**

*Online product reviews are a major source of business intelligence (BI) that helps managers and market researchers make important decisions on product development and promotion. However, the large volume of online product review data creates significant information overload problems, making it difficult to analyze users' concerns. In this paper, we employ a design science paradigm to develop a new framework for designing BI systems that correlate the textual content and the numerical ratings of online product reviews. Based on the framework, we developed a prototype for extracting the relationship between the user ratings and their textual comments posted on Amazon.com's Web site. Two data mining algorithms were implemented to extract automatically decision rules that guide the understanding of the relationship. We report on experimental results of using the prototype to extract rules from online reviews of three products and discuss the managerial implications.*

**Keywords:** Business Intelligence, Data Mining, Knowledge Management, Design Science, Text Mining, Automatic Rule Induction.

## **Introduction**

As electronic commerce supports higher interactivity among customers, user-generated content posted on e-commerce Web sites is growing significantly in recent years. The current Internet (Web 2.0) has evolved from its previous generation so that users are not only consumers of the Web content, but also are producers, oftentimes generating voluminous data of their participation. Online product reviews have become a major source of business intelligence that helps managers and marketers to understand customers. According to management scholar Peter Drucker, “what is value to the customer” may be the most important question to answer in order to realize a business’s mission and purpose (p.65, Drucker 2003). However, the large volume of online product review data creates significant information overload problems (Bowman et al. 1994), making it difficult to analyze customers’ 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. Therefore, an important task of a manager is 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. However, such analysis is time-consuming and does not scale up to the rapidly growing online reviews.

Previous research on analyzing online reviews tries to recommend products, to calculate the utility of the reviews, to identify key product features, to detect false reviews, and to summarize review content (Aciar et al. 2007; Ding et al. 2007; Jindal et al. 2007; Zhang 2008; Zhuang et al. 2006). Most of these works focus on sentiment extraction, review classification, and product recommendation. While data mining techniques are widely used, there is a lack of research on using automatic rule-induction techniques in analyzing these reviews. In addition, the problem of how the textual content of the reviews contributes to the numerical ratings is not widely addressed. Answers to such problem would help marketers and managers to understand the reasons behind certain ratings and hence reveal business intelligence (BI) that supports appropriate actions to address the concerns.

In this paper, we employ a design science paradigm (Hevner et al. 2004) to develop a framework for designing BI systems that correlate the textual content and the numerical ratings of online product reviews. Based on the framework, we developed a prototype for extracting the relationship between the user ratings and their textual comments posted on Amazon.com’s Web site. Two data mining algorithms were implemented to extract automatically decision rules that guide the understanding of the relationship. We report on experimental results of using the prototype to extract rules from online reviews of three products and discuss the managerial implications.

## **Related Works**

### ***Business Intelligence Research***

The term “business intelligence” (BI) is defined as the acquisition, interpretation, collation, analysis, and exploitation of information in business (Chung et al. 2005). BI systems enable organizations to understand their internal and external environments. To support understanding of internal data, one class of BI systems (Carvalho et al. 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 et al. 2009a; 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 Research in BI***

Design science is concerned with the creation and evaluation of new IT artifacts with a goal of meeting business needs (Hevner et al. 2004). Hevner et al. provided seven guidelines for design-science research (Hevner et al. 2004): 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. A recent call for papers from a *MIS Quarterly* special issue on BI research highlights this need. It especially states that “information systems research based on design science 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 et al. 2009). To address the needs of analyzing online textual reviews, we believe technologies in data/text/Web mining can provide valuable input to the design of new IT artifacts relevant to BI. We review below these technologies applied to online product review analysis.

### ***Data/text/Web Mining***

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 et al. 2009b), 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 et al. 2000). These techniques are applied to online user-generated content analysis and to knowledge management.

### **Online User-Generated Content Analysis**

Data/text/Web mining techniques have been applied to analyzing online user-generated content, most notably customer reviews. These applications try to analyze users’ textual reviews to help recommend products, to classify sentiment, to calculate the utility of the reviews, to identify key product features, to detect false reviews, and to summarize review content. For example, Yan et al. developed a dictionary-based method to represent review textual features and used machine-learning techniques to classify the review sentiment (Dang et al. 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 et al. 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 et al. 2007). Besides, product recommendation was done through mapping automatically each sentence of a review into an ontology, which is a manually-created ontology for a product (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 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 et al. 2001)).

## Knowledge Management and Automatic Rule Induction

Knowledge management, an emerging driver of competitiveness in today’s business, helps organizations to fully exploit their knowledge-based resources, hence achieving higher efficiency and effectiveness (Alavi et al. 2001). The role of KM in analyzing online product reviews is to reveal the underlying knowledge of the participants, thus supporting the BI community to capture, reuse, exploit, and share the knowledge. KM can help establish routines for identifying knowledge as well as the experts who possess the knowledge. This crucial knowledge identification aims to define, locate, characterize, and classify the knowledge that is important in organizations. Some of these tasks can be facilitated by using automatic rule-induction methods. A promising approach is called Rough Set Theory (RST) (Pawlak 1982) that is a mathematical approach to identifying decision rules from ambiguous and uncertain data, ill-defined problems, indiscernible relations and classifications, and interdependent attributes. The major advantage of RST over standard statistical techniques is the capability to handle qualitative data (Heckerman et al. 1997; Simoudis et al. 1996). RST has been applied to many domains, such as fault diagnosis (Zhang et al. 2009), interval data clustering (Doumpos et al. 2009), and supply chain management (Gaudreault et al. 2009). Despite the widespread applications, it is surprising to find no prior work on applying RST to online product review analysis.

## A Business Intelligence Extraction Framework

As BI research grows in recent years, many efforts have been put in analyzing and understanding online user-generated content. Online product review is one of the most important types of such content. Despite prior works in analyzing these reviews, identifying rules and patterns – a basic form of knowledge representation – is not widely studied. In this regard, Rough Set Theory (RST) has widespread application in many domains. Surprisingly, no prior work has been found on using RST in online product review analysis. We therefore developed an RST-based framework to address the research needs. In the following, we describe a BI extraction framework and apply it to developing a prototype for analyzing online product reviews.

### Proposed Framework

The various components we consider in developing our framework are shown in Figure 1. A new IT artifact, the framework was developed to enrich the knowledge based 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  $n$  => 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

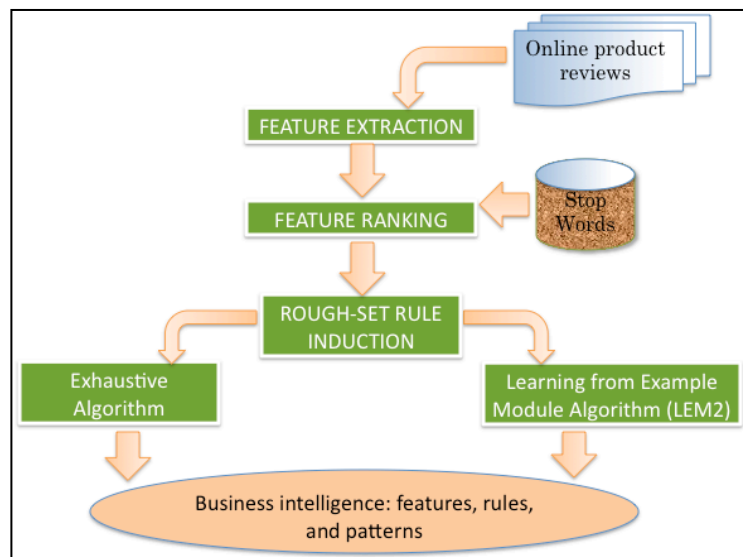


Figure 1. A framework for automated rule-induction from online reviews

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.

**Feature Extraction and Ranking**

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 collection of online reviews 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 adapted from the information retrieval domain to compute term importance in document classification (Salton 1989).

**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  
 $tf_{ij}$  = number of occurrence of term  $j$  in review  $i$   
 $df_j$  = number of reviews containing term  $j$   
 $N$  = number of reviews for a product  
 $n$  = number of unique textual terms appearing in all the reviews

Intuitively, 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 use the term importance scores to rank all extracted features of a product’s online reviews and select the top-ranked features to serve as input of the next steps.

**Rule-induction Algorithms**

RST is concerned with measuring the “ambiguity” inherent in the data. Essential distinction is made between objects that may definitely be classified to a certain category, and those that may possibly be classified. Considering all decision classes with yield to what is referred to as the “quality of approximation,” that measures the proportion of all objects for which definite classification may be achieved. A rough set is a collection of objects that, in general, cannot be precisely characterized in terms of their 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).

**Table 1. Example Data Set**

Review	F <sub>1</sub>	F <sub>2</sub>	F <sub>3</sub>	F <sub>4</sub>	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

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 (x)(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., (x)(F1 = 0) ∧ (F2 = 1) = {1, 3} and (x)(F1 = 0) ∨ (F2 = 1) = {1, 2, 3, 5}, where ∧ and ∨ refers to “and” and “or” respectively. Most of RST-based approaches may generate for an object more than one reduct. Through enumerating all possible reducts, those reducts can be manipulated to produce different decision rules. In this paper, two rough set based algorithms have been used to derive the decision rules. One is called the exhaustive algorithm (Bazan et al. 2000) while the other is named LEM2 (Learning from Examples Module) approach

(Grzymała-Busse 1997). Due to their effectiveness and high prediction accuracy, the exhaustive and LEM2 algorithms are used in our framework.

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—in most cases—it gives best results in terms of accuracy. We are interested in answering these research questions.

1. How can a framework for automatic extraction of decision rules from online reviews be designed and developed?
2. How can an IT artifact developed based on the framework address the challenges of online product review analysis?
3. What is the performance of the IT artifact in inducing decision rules from online product reviews?

## Experimental Findings

To study the feasibility of our proposed framework, we developed a research prototype that implements the two algorithms. Our hypotheses in the experiment are that the framework would successfully guide the development of an IT artifact to perform BI extraction in the form of decision rules, that the artifact would be applicable to real-world data and would achieve satisfactory performance compared with standards accepted in academic literature. We 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 wide range of products. Each review includes a title, a text review, date, time, author name and location, ratings, and other miscellaneous information. We have tested our prototype using the online reviews of three products: 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). These three items represent some of the most popular product categories, namely, jewelry, action sports, and electronics. Table 2 shows the basic information about the reviews. The sizes of these data sets are adequate because the proposed algorithms do not require large data sets for evaluation (i.e., training and testing) (Grzymała-Busse 1997; Nguyen et al. 2003) and it is a similar case as in (Kusiak et al. 2000; Zaki et al. 2000). For our future work, we plan to use larger datasets to test further the performance of our framework.

**Table 2. Statistics of the Product Reviews**

Product	Total no. of reviews	5-star	4-star	3-star	2-star	1-star
Sterling Silver Marcasite & Garnet Glass Heart Pendant 18"	78	40	24	5	4	5
Razor Power Wing Caster Scooter	78	63	10	2	0	3
HP 2133-KR922UT 8.9-Inch Mini-Note PC	123	37	39	13	12	22

### Rule Induction

In the feature extraction, ranking, and filtering processes, we selected as features the top 100 words with the highest term importance scores. This number of terms was selected so that the words represent a small fraction of all distinct words extracted from the three products' reviews (924, 1005, and 3079 words were extracted respectively from the three products listed in Table 2). These features were used as input to the two algorithms, which induced the decision rules using a process based on RST. For instance, in deducing the rule “‘great’  $\rightarrow$  rating = 5-star” from the reviews of razor scooter using the exhaustive algorithm, the reducts from the razor scooter data were first exhaustively generated. Then, the algorithm combined similar reducts (if any) from each data set. Next, the reducts with high frequency support formed the final rules. For example, the rule “‘great’  $\rightarrow$  rating = 5-star” was selected after the final reducts evaluation from the razor scooter data. Listed below are a fragment of the rules derived using the exhaustive algorithm and the LEM2. Each rule is significant if it is supported by at least two reviews. This number was determined by aggregating the numbers provided by a panel of domain experts in online product reviews.

Exhaustive algorithm result

Rule ID	Frequency	Rule
1	11	(F12=1)=>(rating={4(4), 5(5), 3(1), 1(1)})
2	2	(F2=2)&(F26=1)&(F66=2)=>(rating={5(1), 4(1)})
3	2	(F8=1)&(F57=2)&(F65=1)=>(rating={4(1), 5(1)})
4	2	(F330=2)=>(rating={2(2)})
LEM2 algorithm		
1	3	(F65=1)&(F26=1)&(F17=1)=>(rating={5(3)})
2	3	(F2=1)&(F57=2)=>(rating={5(3)})
3	3	(F2=1)&(F9=1)&(F26=1)=>(rating={4(3)})
4	3	(F8=1)&(F17=1)&(F26=1)=>(rating={4(3)})

In the first rule (F12=1)=>(rating={4(4),5(5),3(1),1(1)}), it means when the word “Piece” is present, there are 4 reviews with a 4-star rating, 5 reviews with a 5-star rating, etc.. Since 5-star obtains more supports, it is a very strong rule. Consequently, the rule of rating = 1-star is very weak and therefore it can be neglected. In the 2nd rule derived by the LEM2 algorithm, (F2=1) & (F57=2)=>(rating={5(3)}), this means when “My” = 1 and “Beautiful” = 2, 3 objects support rating = 5-star. Note that this a compound rule different from the previous rule. Table 3 illustrates number of rules based on different approaches and with or without filtering operation. Note that the value inside the parentheses indicates the highest frequency (i.e., number of reviews supporting the rule) associated with in each rule sets.

**Table 3. Summary of rule sets derived from the exhaustive and LEM2 algorithms**

Algorithm	Sterling Silver Marcasite	Razor Scooter	HP Mini-Note PC
Exhaustive	7 (5)	30 (26)	6 (3)
After filter	4 (5)	27 (26)	2 (3)
LEM2	23 (3)	21 (9)	34 (3)
After filter	23 (3)	21 (9)	34 (3)

Table 3 shows the number of rules induced from each of the three sets of product reviews in our research test bed. Since the rule set for the Razor Scooter contains the highest frequency of the rules when we compare with each sample size, the rules generated from this data set are selected for the following empirical validation (i.e., semantic appraisal) to find out if the rules make sense or not.

In our analysis of the terms used in the reviews of Razor Scooter, we found a number of interesting rules as follows. We underlined the features identified by the algorithms.

Rules induced by the Exhaustive Algorithm

(1) “great” → rating = 5-star

“**Great!** – We bought this for our 11 years old son and now everybody want to play with it. It is a good ride.”

(2) “fun” and “my” → rating = 5-star

“Loves It! – “We bought this item for our 6 year old son for Christmas. It has, by far, been his favorite gift. Everyone that sees it wants to try it out and they all find it just as much **fun** as **my** son!”

“Lots of **fun!** **My** son is 8 and loves scooters. The Power Wing is one of his favorites. This scooter is easy to balance and turns fast. Good quality and easy to put together.”

(3) “my” and “great” → rating = 5-star

“... Easy to assemble, so she rode it on xmas day! As with all the razor products, this powerwing is built to last..solid..which is what is needed for kids!! A **great** item!!”

Rules induced by the LEM2 Algorithm

(1) “year” and “great” → rating = 5-star

“**GREAT** POWER WING We got this for my 5 **year** old son for Christmas. He loves it and could ride it the first day with no problems.”

(2) “fun,” “my,” and “easy” → rating = 5-star



“My son likes it! "My 11 year old son got this for **Christmas**. He has only used it once because it has been so snowy, but he really likes it. It looks like **fun**. I wish I could use it!”

(3) “bought,” “my,” “kids,” and “great” → rating = 5-star

“This is a really **great** scooter for both boys and girls. The retailer I **bought** this from was wonderful and got it here just in time for **my** 9 year old Granddaughter's Christmas surprise... It's easy for the **kids** to operate and control.”

**Rule Validation**

To validate the quality of the induced rules, we applied a 10-fold cross-validation which is a recommended scheme for rule evaluation (Stone 1974). In a 10-fold cross-validation scheme, the entire data set is partitioned randomly into 10 folds (disjoint subsets). One fold is removed from the data set to serve as testing data while the remaining nine folds serve as training data from which rules are extracted. Then, decision ratings (i.e., outcomes of rules) are predicted for each review in the testing data based on the rules extracted from the training data. The performance metrics (accuracy and coverage, explained below) are computed. Next, another fold of data serves as the testing data while the remaining nine folds serve as training data. This process is repeated systematically for 10 times where each of the 10 folds serves once as testing data in each validation. The ten sets of performance figures are then averaged to provide overall performance metrics.

We used accuracy and coverage as our performance metrics, which are widely-accepted metrics in evaluating RST methods (Skowron 2005; Tsumoto 2002). Accuracy is the ratio of correctly classified reviews from the class to the number of all reviews assigned to the class by the classifier. For instance, if 210 reviews (out of all the 216 reviews with 5-star ratings) are classified correctly as 5-star reviews, then the accuracy is  $210/216 = 0.972$ . The accuracies in multi-class classification are averaged to give an overall accuracy. Coverage is the ratio of classified (recognized by classifier) reviews from the class to the number of all reviews in the class (i.e., percentage of test objects that were recognized by classifier). In our example, the total coverage equals 1, which means that all reviews have been recognized (classified). Such total coverage is not always the case, as the constructed classifier may possibly not recognize previously unseen review. If some reviews remain unclassified, the total coverage is less than 1. The performance of each algorithm in terms of accuracy and coverage is listed in Table 4.

**Table 4. Accuracy and coverage of the two algorithms**

Product	Sterling Silver Marcasite		Razor Scooter		HP Mini-Note PC	
	Accuracy	Coverage	Accuracy	Coverage	Accuracy	Coverage
The Exhaustive Algorithm	0.329	1	0.529	1	0.258	1
LEM2	0.4	0.614	0.623	0.614	0.294	0.475

Comparing performance of the two proposed algorithms, we conclude that the exhaustive algorithm is excellent in total coverage because it was able to induce rules from all the reviews in each of the three data sets. The accuracy of LEM2 is very good because it is higher than the accuracy of the exhaustive algorithm in each of the three data sets and it is significantly higher than the accuracy that a random classification would obtain (i.e., 0.2 due to this five-class prediction).

**Managerial Implication of the Induced Rules**

To explicate the managerial implication, we describe below a hypothetical case that demonstrates the use of the induced rules by the BI manager of a major online retailer. The manager was used to conducting manual analysis of online reviews. In such analysis, he read 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 a 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.

## Conclusion and Future Works

To understand what the customer values in a product, managers often analyze online product reviews to correlate between the numerical ratings and the textual content. However, the large volume of online reviews makes it difficult for managers to perform this analysis manually. To support automatic BI extraction from these online reviews, we developed and validated a new framework for designing BI systems that extract rules and patterns from textual content of the reviews. The framework and the BI system developed based on it are two new IT artifacts that promise to contribute to knowledge management by extracting business intelligence from online reviews. Two rule-induction data mining algorithms were used to extract automatically decision rules that guide the understanding of the relationship between the terms used in online reviews and the product ratings. Interesting rules and patterns were identified. The exhaustive algorithm produced excellent coverage while the LEM2 algorithm produced very high accuracy. These IT artifacts provide new tools to managers and marketers to analyze their rapidly-growing online product reviews.

Our ongoing work includes expanding the review datasets to cover more products, testing other data mining algorithms for knowledge extraction, and studying reviews with a majority of low evaluations. This work should contribute to advancing the understanding of user-generated content analysis, data and text mining, and business intelligence technology because of the following reasons. The online reviews we study belong to a major type of user-generated content. Our framework incorporating data and text mining techniques demonstrates a novel application of these techniques to user-generated content (online reviews). Because online reviews are perceived to be more important than personal recommendation, the new IT artifacts (the framework and the BI system) developed in this research contribute to the domain of BI technology.

## Acknowledgments

This paper is based upon work supported partially by the Santa Clara University Presidential Research Grant (2009-2010) and by the National Science Foundation (through award CCF-0752865 and its REU supplement). Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of NSF. We thank the ICIS track chairs and reviewers for their valuable comments and suggestions.

## References

- Aciar, S., Zhang, D., Simoff, S., and Debenham, J. "Informed recommender: basing recommendations on consumer product reviews," *IEEE Intelligent Systems* (22:3) 2007, pp 39-47.
- Alavi, M., and Leidner, D.E. "Review: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues," *MIS Quarterly* (25:1) 2001, pp 107-136.
- Bazan, J., Nguyen, H.S., Nguyen, S.H., Synak, P., and Wróblewski, J. "Rough set algorithms in classification problem," in: *Rough Set Methods and Applications*, L. Polkowski, S. Tsumoto and T. Lin (eds.), Physica-Verlag Heidelberg New York, 2000, pp. 49-88.

- Bowman, C.M., Danzig, P.B., Manber, U., and Schwartz, F. "Scalable Internet resource discovery: research problems and approaches," *Communications of the ACM* (37:8) 1994, pp 98-107.
- Brown, M.P., Grundy, W.N., Lin, D., Cristianini, N., Sugnet, C.W., Furey, T.S., Ares, M., and Haussler, D. "Knowledge-based analysis of microarray gene expression data by using Support Vector Machines," *Proceedings of the National Academy of Science*, 2000, pp. 262-267.
- Carvalho, R., and Ferreira, M. "Using information technology to support knowledge conversion processes," *Information Research* (7:1) 2001.
- Chen, H. *Knowledge Management Systems: A Text Mining Perspective* The University of Arizona, Tucson, AZ, 2001.
- Chen, H., Chiang, R.H.L., and Storey, V.C. "MIS Quarterly Special Issue on Business Intelligence Research," in: *MIS Central*, <http://www.misq.org/>, 2009.
- Chen, H., Chung, W., Xu, J.J., Wang, G., Chau, M., and Qin, Y. "Crime data mining: a general framework and some examples," *IEEE Computer* (37:4) 2004, pp 50-56.
- Chung, W. "An Automatic Text Mining Framework for Knowledge Discovery on the Web," in: *Department of Management Information Systems*, The University of Arizona, Tucson, AZ, 2004, p. 270.
- Chung, W. "Visualizing E-Business Stakeholders on the Web: A Methodology and Experimental Results," *International Journal of Electronic Business* (6:1) 2008, pp 25-46.
- Chung, W., and Chen, H. "Web-based Business Intelligence Systems: A Review and Case Studies," in: *Handbooks in Information Systems: Business Computing*, G. Adomavicius and A. Gupta (eds.), Emerald Group Publishing Limited, 2009a, pp. 373-396.
- Chung, W., Chen, H., and Nunamaker, J.F. "A visual framework for knowledge discovery on the Web: An empirical study on business intelligence exploration," *Journal of Management Information Systems* (21:4) 2005, pp 57-84.
- Chung, W., Chen, H., and Reid, E. "Business stakeholder analyzer: An experiment of classifying stakeholders on the Web," *Journal of the American Society for Information and Science and Technology* (60:1) 2009b, pp 59-74.
- Dang, Y., Zhang, Y., and Chen, H. "A lexicon-enhanced method for sentiment classification: An experiment on online product reviews," *IEEE Intelligent Systems* (25:4) 2010, pp 46-53.
- Ding, X., and Liu, B. "The utility of linguistic rules in opinion mining," *Proceedings of the 30th annual international ACM SIGIR Conference on Research and Development in Information Retrieval*, ACM New York, NY, USA, Amsterdam, The Netherlands, 2007, pp. 811-812.
- Doumpos, M., Marinakis, Y., Marinaki, M., and Zopounidis, C. "An evolutionary approach to construction of outranking models for multicriteria classification: The case of the ELECTRE TRI method," *European Journal of Operational Research* (199:2) 2009, pp 496-505.
- Drucker, P. *The Daily Drucker* Harper Collins, New York, NY, 2003.
- Gaudreault, J., Frayret, J.M., and Pesant, G. "Distributed search for supply chain coordination," *Computers in Industry* (60:6) 2009, pp 441-451.
- Grzymała-Busse, J. "A New Version of the Rule Induction System LERS," *Fundamenta Informaticae* (31:1) 1997, pp 27-39.
- Heckerman, D., Mannila, H., Pregibon, D., and Uthurusamy, R. *Proceedings of the Third International Conference on Knowledge Discovery and Data Mining*, Menlo Park AAAI Press, CA, 1997.
- Hevner, A.R., March, S.T., Park, J., and Ram, S. "Design science in information systems research," *Management Information Systems Quarterly* (28:1) 2004, pp 75-105.
- Jindal, N., and Liu, B. "Analyzing and detecting review spam," *Proceedings of the 7th IEEE International Conference on Data Mining*, IEEE Press, Omaha, NE, 2007, pp. 547-552.
- Kosala, R., and Blockeel, H. "Web mining research: a survey," *ACM SIGKDD Explorations* (2:1) 2000, pp 1-15.
- Kusiak, A., Kern, J.A., Kernstine, K.H., and Tseng, B. "Autonomous Decision-Making: A Data Mining Approach," *IEEE Transactions on Information Technology in Biomedicine* (4:4) 2000, pp 274-284.
- Lawton, G. "Making business intelligence more useful," *IEEE Computer* (39:9) 2006, pp 14-16.
- Liu, B. *Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data* Springer-Verlag Berlin Heidelberg, 2007.
- Mitchell, T.M. *Machine Learning* McGraw-Hill, New York, 1997.
- Nguyen, S.H., and Nguyen, H.S. "Analysis of STULONG Data by Rough Set Exploration System (RSES)," *Proceedings of The European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD) Discovery Challenge Workshop*, Springer, Cavtat, Croatia, 2003, pp. 71-82.

- pattern," in: *Merriam-Webster Online Dictionary*, <http://www.merriam-webster.com/dictionary/pattern>, retrieved on August 14, 2010.
- Pawlak, Z. "Rough sets," *International journal of computer and information sciences* (11:5) 1982, pp 341-356.
- Salton, G. *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer* Addison-Wesley, Reading, MA, 1989.
- Simoudis, E., Han, J., and Fayyad, U. *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, Menlo Park, CA* AAAI Press, 1996.
- Skowron, A. "RSES 2.2 User's Guide," Warsaw University, [http://logic.mimuw.edu.pl/~rses/RSES\\_doc\\_eng.pdf](http://logic.mimuw.edu.pl/~rses/RSES_doc_eng.pdf) 2005.
- Stone, M. "Cross-validators choice and assessment of statistical predictions," *Journal of the Royal Statistical Society. Series B (Methodological)* (36:2) 1974, pp 111-147.
- Trybula, W.J. "Text mining," in: *Annual Review of Information Science and Technology*, M.E. Williams (ed.), Information Today, Inc., Medford, NJ, 1999, pp. 385-419.
- Tsumoto, S. "Accuracy and Coverage in Rough Set Rule Induction," in: *Rough Sets and Current Trends in Computing, Lecture Notes in Computer Science* Springer, Berlin / Heidelberg, 2002.
- Zaki, M.J., and Ho, C.-T. (eds.) *Large-Scale Parallel Data Mining*. Springer-Verlag, New York, 2000.
- Zhang, Z. "Weighing stars: Aggregating online product reviews for intelligent e-commerce applications," *IEEE Intelligent Systems* (23:5) 2008, pp 42-49.
- Zhang, Z., Shi, Y., and Gao, G. "A rough set-based multiple criteria linear programming approach for the medical diagnosis and prognosis," *Expert Systems with Applications* (36:5) 2009, pp 8932-8937.
- Zhuang, L., Jing, F., and Zhu, X.-Y. "Movie review mining and summarization," *Proceedings of the 15th ACM international conference on Information and knowledge management*, ACM New York, NY, USA, Arlington, Virginia, USA, 2006, pp. 43-50.