

Editorial: Data Mining in Electronic Commerce – Support vs. Confidence

César Astudillo¹, Matthew Bardeen² and Narciso Cerpa³

Universidad de Talca, Faculty of Engineering, Curicó, Chile

¹castudillo@utalca.cl, ²mbardeen@utalca.cl

³Editor-in-Chief

January 2014

Introduction to Data Mining and Electronic Commerce

In the year 2001, one of the authors of this editorial wrote an article about *support versus confidence* in the data mining technique, association rules. This article was presented at a conference and never formally published [22]. In the last four years this article has been downloaded nearly twenty-thousand times from an open access repository. This interest by researchers and practitioners has motivated us to write this technical editorial. The structure of this editorial will be as follows. In this section we briefly introduce data mining and electronic commerce. In the following section we describe different data mining techniques. In the final section we discuss the effect of support versus confidence in association rules technique applied to electronic commerce.

The data mining process involves searching, selecting, exploring, and modeling large amounts of data to uncover previously unknown patterns that are potentially useful, and ultimately comprehensible information, from large databases. Its goal is to manipulate data into knowledge ([15], [18]-[19], [31], [33]). Pattern extraction is an important process of any data mining technique and it refers to the relationships between subsets of data.

Data mining use different families of computational, statistical and machine learning methods that include statistical analysis, decision trees, neural networks, rule induction and refinement, and graphic visualization among others, to exhaustively explore data to reveal complex relationships that may exist. Although machine learning techniques have been available for a long time, the development of advanced and user friendly tools for business intelligence [25] has made data mining more attractive and practical for organizations. When these pattern extraction techniques are used correctly, they can be effective tools for extracting useful information from data [35].

The recent wide use of data mining has been due to several factors. The most obvious of these is the large amounts of data that organizations collect during operational transactions. In the early 90s, credit and insurance companies began using data mining as a means of detecting fraud [28]. Most organizations, irrespective of the industry type, have some form of operational process in which they collect large amounts of data. For example, the retail industry has been using data mining techniques for years to predict what their customers are likely to purchase. The electronic commerce industry was one of the latest to use data mining technology [18].

Electronic commerce is the use of information and communication technologies through the Internet platform to share business information, keep business relationships, and conduct business transactions. In electronic commerce, different data mining techniques can be used for many purposes. For example, in sales promotion the marketing staff may want to find out which products their customers are more likely to buy together. This information will allow them to place these items in a sales bundle in order to increase revenue ([2], [31]). The use of Web log data permits to understand users' behavior. This data contains information about users' access and may show potential patterns in their behavior, and identify potential customers of electronic commerce. This knowledge is useful to: change marketing strategies; identify segmentation of customers; improve customers' retention; predict customer's expenditure and market trends; provide personalized services to customers; analyze shopping cart; forecast sales; redesign the website to provide a better service; and/or make better business decisions. This area of data mining has given rise to Web mining, a technique that can be subdivided into Web content mining; Web structure mining; and Web usage mining ([7], [24]). These techniques are also used to extract useful information from Web documents or Web services [5] and are widely used in a variety of applications.

As we describe above, data mining and specifically web data mining technology plays an important role in electronic commerce. In recent years with the rapid growth of electronic commerce and the large amounts of data collected through operational transactions, data mining techniques are becoming more useful to discover and understand unknown customer patterns. In the following paragraphs we briefly describe some examples of the application of data mining in electronic commerce.

Clustering or grouping electronic commerce customers with similar browsing behaviors permit the identification of their common characteristics, providing a better understanding of customers with the aim of giving them a more appropriate, and personalized service. When a vendor knows the customer's needs and interests, they can work on providing a better service and keeping the customer relationship with the vendor.

Electronic commerce organizations use web data mining to obtain reliable market and client feedback. For example, the information obtained may help organizations to undertake targeted marketing, decide advertisement positioning, and reduce operating costs.

The appropriate understanding of customer behavior and feedback also helps software designers to improve the website design — optimizing its structure and facilitating navigation by customers. Techniques such as association rule mining permit the analysis of shopping cart data to improve the presentation or location of products.

Those in charge of website and electronic commerce security may look at fraud detection, together with financial organizations such banks, and credit card companies, with the aim of detecting fraudulent use of their credit cards. Differences in the customer's spending pattern may suggest a possibility of fraud. In intrusion detection the data mining algorithms may show that a certain sequence of events may indicate that there is an unauthorized access attempt by hackers ([8], [31]-[32]). Understanding these patterns may help computer security personnel to prevent future intrusions.

We can say that the use of data mining in electronic business and electronic commerce has been defined as a data-centric approach called Business Intelligence and Analytics (BI&A) 1.0 [9] which make use of structured content, while the use of web mining to uncovering patterns in web-based and unstructured content has been labeled BI&A 2.0 [13]. Currently, with the introduction of mobile devices a new research opportunity has arisen to analyze mobile and sensor-based content (BI&A 3.0) [10]. Some new areas of research have emerged and existing ones have increased their strength (i.e., Big Data Analytics; Text Analytics; Web Analytics; Network Analytics; and Mobile Analytics). This has given life to more challenging problems [43] and specifically to Big Data Analytics.

The current electronic commerce systems used by the major Internet organizations such as Google, Amazon, Facebook, etc. differ from the systems used initially in this field (shopping cart analysis making use of structured data). They include highly scalable electronic commerce platforms, product recommender systems, social media platforms, and make use of web data that is less structured and that usually is composed of rich customer views and behavioral information [10]. The analysis of customer opinion in social media has used text analysis and sentiment analysis techniques ([17], [30], [36]). Product recommender systems use customer segmentation and clustering, anomaly detection, graph mining, and more importantly association rule techniques [1]. The use of highly targeted searches and personalized recommendations has made possible long-tail marketing to reach millions of small niche markets [3].

Types of Patterns that can be Mined

Machine learning is a mature area of computer science that researches how computers learn patterns and regularities in the data. Data mining, on the other hand, is performed by a human person with a specific goal. Usually, this person utilizes one or more pattern recognition algorithms that have been created in the machine learning field. This person deals with situations in which the data is available in impressive amounts and which probably possess some deficiencies such as missing data or high dimensionality when compared to the cardinality of the observation set.

Data mining can be organized according to the different family of problems that it solves. These problems include classifying items into previously known categories, grouping items according to their similarities, discovering association rules from transactions, identifying atypical data, and predicting a continuous (dependent) variable. This section will give a brief overview of these types of problems.

Classification

In data mining applications one often assumes that the data is already in some sort of digital form (something like a big spreadsheet). Here, one may want to predict the value of a particular attribute (a particular column in the spreadsheet). When this attribute, sometimes referred to as the class attribute, includes a finite number of discrete elements, we are in presence of a classification problem. In this type of problem, we build a mathematical model from the available data. This model receives the information of a novel instance whose class is unknown and produces an estimation of the category which it belongs to. Our task is to perform this estimation as accurately as possible.

In machine learning, classification is a form of supervised learning where instances or items are assigned to some predefined set of categories $\Omega = \{\Omega_1, \Omega_2, \dots, \Omega_C\}$. More formally, data classification is a mathematical function g that is constructed from a collection of instances $\mathcal{X}_L = \{x_1, x_2, \dots, x_n\}$ called the training set. In classification, the class memberships of all the training instances in the category are known in advance. The categories of the instances are contained in a vector $\mathbf{y} = \{y_1, y_2, \dots, y_n\}$, where the category of instance i is denoted by $y_i \in \Omega$. The fundamental idea behind classification, is that there is an underlying function f that relates the patterns and their respective categories. Unfortunately f is unknown to us and we want to estimate it by building a function g from the patterns and its categories, following the procedure specified by a learning algorithm \mathcal{A}_S .

Recent applications of data classification include social network classification [38], credit scoring [37], fraud detection [21], web mining to predict e-commerce company success [34], among many more.

Clustering

Sometimes we desire to categorize elements even when the set of categories Ω are not available. This problem is known as data clustering and represents a more challenging task from a learning perspective when compared to data classification. Here, our mathematical model receives the data without the class labels and we expect it to infer groups of elements just by merely examining their similarities. The output is an estimated class membership. In contrast to the classification problem where there is a set of possible classes known *a priori*, in the clustering problem different groups are created. The objective is to group similar instances in the same group, while at the same time, assign to distinct groups those elements which are different. This type of learning is sometimes referred to as unsupervised learning because the function lacks of a teacher that tells the correct class label of a particular pattern. Formally, data clustering, or sometimes referred to as hard clustering [20] consist in assigning a class label l_i to each pattern x_i in the set $X_U = \{x_1, x_2, \dots, x_n\}$, identifying its respective label. The set of all labels for a pattern set X_U is $L = \{l_1, l_2, \dots, l_n\}$, where $l_i \in \{1, 2, \dots, k\}$, and where k is the number of clusters.

Interestingly enough, the human brain is particularly good at this task. For generations we have used this type of reasoning to distinguish between ripe fruit before collecting it or to build an entire taxonomy of the animal kingdom based on the observed characteristics: Nobody told us to categorize animals according to whether they produce milk or not!

Applications of data clustering in e-commerce include recommendation systems [40], search engines [23], etc.

Semi-supervised Classification

Classification is an example of supervised learning, assuming the knowledge of well-defined training sets with a clear specification of the identity of *all* the training samples. A distinct and intriguing learning paradigm that has emerged in the recent years is semi-supervised learning. This paradigm combines labeled and unlabeled instances simultaneously to perform classification [46]. This specific type of classifiers does not demand the specification of the class labels of every sample. Usually this type of learning appears in situations where many instances are available, but only few of them possess labels because the cost of acquiring them is high. One common way to learn in this context is to perform a clustering-like mechanism, assigning the training samples into different groups, and subsequently, a class label is assigned to each group using a small subset of the training instances whose class identities are known. Given a clustering algorithm, \mathcal{A}_C , a set of labeled instances, X_L , a set of unlabeled instances, X_U , and a supervised learning algorithm, \mathcal{A}_S , the Cluster-then-Label method works as follows [4]: First, we identify the clusters of the input manifold using the clustering algorithm \mathcal{A}_C . Secondly, we determine which of the labeled samples fall in each cluster. For each cluster we determine a decision boundary based on the supervised algorithm \mathcal{A}_S , and the labeled samples assigned to that cluster, which, in turn, allows the prediction of the label of every cluster. Finally, each uncategorized item is labeled according to the predicted class of the cluster in which it is contained. Recently, semi-supervised classification has been successfully applied in the estimation of the quality of online reviews [45].

Association Analysis

Another major problem studied in data mining is the so-called association analysis. In this context, our data is conformed of transactions, e.g., the bill that includes a list of products that you bought in a grocery store. The nature of the data is unique: items do not necessarily repeat in two bills, but usually people tend to behave similarly in their buying trends. Association analysis attempts to discover those trends. In this context, one may be interested in knowing which patterns are more frequent. A famous example is the relationship between diapers and beer in grocery store bills. Information like this provides useful information at the time a grocery store is designed: if you know that people will buy beer and diapers you can put them together, or place them in opposite corners, increasing the probability that the customers will see other products they might be interested in!

Association rule mining can be formally defined as follows [2]: Let $I = \{a_1, a_2, \dots, a_n\}$ be a collection of n elements called items. Also, let $D = \{T_1, T_2, \dots, T_m\}$ be a collection of transactions called the database. Each transaction $T \in D$ contains a subset of the items in I . Additionally, an itemset is a set of items. Given an itemset $X \subseteq I$ and a given transaction T , it is said that T contains X if and only if $X \subseteq T$. The support count of a given itemset X , denoted by σ_X , is defined as the number of transactions in D that contain X . Let s be the support threshold and $|D|$ be the total number of transactions in D . An itemset is said to be frequent if $\sigma_X \geq D \times s\%$. An association rule corresponds to an implication $X \Rightarrow Y$ where $X, Y \subseteq I$, and $X \cap Y = \emptyset$. One of the main goals of association analysis is to discover association rules or sets of frequent items.

Applications of association analysis include customer relationship management (CRM) [39], building recommendation systems [44], personalization applications and collaborative filtering [27].

Regression Analysis

In classification we are predicting a discrete variable y , whereas in regression this output variable is of a continuous nature. Let \mathcal{X} be the input space and \mathcal{Y} be a measurable subset of the real domain. Let D be a collection of observations, where there is an underlying function $f: \mathcal{X} \rightarrow \mathcal{Y}$ that relates the observation x with a dependent variable y . In the regression problem f is assumed to be unknown, and the goal is to build a function g that approximates f as closely as possible. Analogous to classification, regression is an example of supervised learning because the learning model receives n pairs of the form (x, y) where x is the observation and $y = f(x)$ corresponds to the label. Since in regression the output is continuous, it is not realistic to expect that the learning model exactly predicts the output of a given instance. Instead, the performance of the regression model is usually measured by computing the difference between the predicted value and the correct one. This is a key difference in respect to classification, where the class labels are expected to be identical or not. Formally, the performance of the regression model is computed with a loss function $L: \mathcal{Y} \times \mathcal{Y} \rightarrow \mathbb{R}$. According to the authors of [26], the most common loss function is $L_2 = L_{y, \hat{y}} = |y - \hat{y}|^2$, where y is the true label and \hat{y} is the one that the model predicted. Regression analysis has been applied to many areas of electronic commerce such as electronic tourism [29], mobile commerce [11], etc.

Outlier/Anomaly Detection

Outlier detection or anomaly detection is another problem of interest in data mining. This problem deals with the identification of patterns that are very different to the rest of the exemplars. Outlier detection appears in security, which is one of the most important areas of e-commerce: Users must trust in a company's website before they agree to enter their credit card details. One major area in security is intrusion detection [6].

Most Prominent Data Mining Algorithms

In 2006, in an effort to identify the most influential papers in the field, the organizers of the IEEE International Conference on Data Mining performed a brief survey of the participants. As a result, a paper was published in the Journal Knowledge and Information Systems [41] specifying ten algorithms: C4.5, k-means, SVM, Apriori, EM, Page Rank, Adaboost, KNN, Naive Bayes and CART. Knowing how to apply these methods in electronic commerce will be a great start for performing data mining.

Remarks

Data mining plays an important role in the development of the electronic commerce. Moreover, machine learning is key in data mining since it provides several algorithms capable of learning from data. We have described different types of patterns that can be mined including classification, clustering, anomaly detection, regression, association analysis and semi-supervised learning. There is no best algorithm known for any of the sub-problems described in this section. This leaves an important task to the data miner, who has to devise manual or semi-automatic methods for identifying the most suitable algorithm for a particular application, and also refining its parameters and interpreting the results.

Association Rules and Association Analysis

We have described different data mining techniques in the previous section, but we now wish to focus on just one commonly used method — association rule mining. As mentioned previously, association rule mining is typically used to discover unknown relations that exist within data. To mine association rules successfully, the data miner must choose which type of rules to mine — are novel and/or unique rules more important or are common rules more important? The answer to this question will determine the method used to mine the rules. We will now illustrate the ramifications through an example and subsequent discussion of *support* vs. *confidence* in association rule mining.

One common application of this technique is to discover relations between items in a shopping basket. For example, if milk and ice-cream are commonly bought together, this could be discovered by using association rule mining algorithms. To discover these relations, we must examine the co-occurrence of items within the same basket, over many baskets. On doing so, we might discover that two items always appear within the same basket. This gives a great degree of *confidence* that there is a relationship between the two items. However, if in our data set these two items only appear once in hundreds of baskets, we do not have a lot to *support* our rule i.e., it could just be a spurious relationship. If the items appear frequently in our list of baskets, then we have a higher support for our rule. To discover these rules there are two main approaches — support-then-confidence and confidence-then-support [22]. In the first, we look for highly supported rules first, then worry about the confidence whereas in the second we do the opposite.

We defined the concept of support previously, but as a reminder, support is defined formally as:

$$s(X \rightarrow Y) = \frac{\sigma(X \cup Y)}{|D|},$$

and confidence is defined as:

$$c X \rightarrow Y = \frac{\sigma X \ Y}{\sigma X}.$$

In our list of example transactions to the right, Milk has a support:

$$\frac{\sigma(\text{Milk})}{5} = \frac{4}{5} = 0.8.$$

Through the same formula, Hot Sauce has a support of 0.2. Mining this data we can find some relationships between the items. For example, $\{\text{Milk} \rightarrow \text{Ice Cream}\}$, and $\{\text{Milk} \rightarrow \text{Bread}\}$ are both well represented in the data.

The confidence of the rule $\{\text{Milk} \rightarrow \text{Ice Cream}\}$ is:

$$\frac{\sigma(\text{Milk, Ice Cream})}{\sigma(\text{Milk})} = \frac{2}{4} = 0.5,$$

and has a support of 0.4. The rule $\{\text{Milk} \rightarrow \text{Bread}\}$ has the same support and confidence, whereas $\{\text{Hot Sauce} \rightarrow \text{Jalapeño Peppers}\}$ has a low support of 0.2 but a confidence of 1.

The principal task for data mining is to discover all of the rules that have some minimum support and some minimum confidence. A brute force approach is infeasible, as the number of possible rules grows exponentially with the number of items in the data set. Worse, the majority of these possible rules will not satisfy reasonable minimal expectations for either support or confidence.

There are a number of approaches that can be used for discovering appropriate rules, but the most basic is the Apriori algorithm. This algorithm works on the assumption that if a particular combination of items is frequent in the dataset, each of the individual items is also frequent. Inversely, if an individual item is infrequent, combinations that include that item will be infrequent as well. This information can be used to *prune* potential rules based on the support that they have in the dataset. This approach can be considered an example of a support-then-confidence rule, where we look at the support an association has first, then consider its confidence.

However, this approach only works if the desired support level is set at high values. If we wish to find infrequent items with high confidence values, the performance of the Apriori algorithm is almost indistinguishable from a brute force method. We also must distinguish between spurious weakly-correlated items, where one item has high support but the other has a much lower support. Most people may buy milk when they shop, but only a few may buy hot sauce — as a result, the chance is high that the same people that buy hot sauce also buy milk, even though this is clearly not a casual relationship [42]. In our example above, $\{\text{Hot Sauce}, \text{Jalapeño Peppers}\}$ has high confidence and low support — yet it likely represents an interesting rule.

The problem of finding items with low support but high confidence is directly related to the *long tail* phenomena in electronic commerce, where there are few items that are purchased frequently, but many more items that are purchased infrequently ([3], [16]). Taking advantage of associations between infrequently purchased items requires that we focus on associations that have low support but high confidence while also filtering spurious correlations. These low support items also relate to finding intrusion events in a log of system accesses or detecting instances of fraud. In each of these cases, instances of intrusions and fraud are likely to be very rare, scattered among many legitimate access and transactions. To find them requires searching for the proverbial needle in the haystack. A good example would be that of fraud — the European Central Bank reports fraud accounted for 18 out of every one hundred thousand credit card transactions in the EU in 2010 [14]. If we were mining this data set, the support of all fraudulent transactions would correspond to just 0.00018.

For these situations, there are better algorithms available than the Apriori algorithm, such as the correlated association rule algorithms [12]. These algorithms do not prune the unsupported rules, but rather look for rules that have high confidence but little support. They are an example of the confidence-then-support approach to association rule mining. Cohen et. al. describe three main stages: first the data is summarized in a hash-signature table, second candidate pairs of items are generated from the summary table, and finally the correlation of these pairs is found using the original data [12]. This approach still faces the combinatorial explosion problem described in the Apriori algorithm. Another approach is that of hyperclique pattern discovery, where an objective measure called h-confidence is used to discover highly correlated patterns in the mining process [42]. This approach uses the h-confidence to prune those item sets that have low confidence and retain those that have high confidence.

These approaches will likely lead to many rules being discovered that are not of much use. In our example above, only one transaction contains the association $\{\text{Hot Sauce}, \text{Jalapeño Peppers}\}$. If we were mining the data for all such rare transactions we would likely find many that fit such low support levels but high levels of confidence — this is the fundamental problem of finding associations in the long tail of many markets. The task then falls to a human expert to weed out the spurious correlations and focus on the rules of interest. For example, perhaps we are looking for

Example Transactions
{Milk, Ice Cream}
{Milk, Bread}
{Milk, Bread, Ice Cream}
{Ice Cream, Bread}
{Milk, Hot Sauce, Jalapeño Peppers}

high value or expensive correlated items. Therefore we must look for rules that only involve those expensive items. In short, the rule mining process must be tailored to the aims of the organization.

Conclusions

In this editorial, we have presented a view of data mining in electronic commerce, and described some of the basic problems that can be solved by the techniques behind this field. We have also provided an overview of some of the more commonly used machine learning techniques in data mining. As an example, we have chosen to describe the association rules mining technique in more detail and have talked about some of the issues associated with its application. In the future, we envision providing descriptions and discussion of other techniques in data mining and how they can be applied to electronic commerce.

In summary, data mining plays an important role in the development of electronic commerce applications. Electronic commerce and the fields related to business intelligence and analytics have developed greatly due to the maturity of data mining and other areas. These areas of research have become more important due to the arrival of big data and the subsequent illumination of the long tail in electronic commerce.

References

- [1] G. Adomavicius and A. Tuzhilin, Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions, *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 6, pp. 734-749, 2005.
- [2] R. Agrawal, T. Imieliński and A. Swami, Mining association rules between sets of items in large databases, in *Proceedings of the ACM SIGMOD International Conference on Management of Data, SIGMOD '93*, New York, 1993, pp. 207-2196.
- [3] C. Anderson. (2004, October). The long tail. *WIRED Magazine*. [Online]. Available: <http://www.wired.com/wired/archive/12.10/tail.html>
- [4] C. A. Astudillo and B. J. Oommen, On achieving semi-supervised pattern recognition by utilizing tree-based SOMs, *Pattern Recognition*, vol. 46, no.1, pp. 293-304, 2013.
- [5] D. L. Banks and Y. H. Said, Data mining in electronic commerce, *Statistical Science*, vol. 21, no. 2, pp. 234-246, 2006.
- [6] H. Bidgoli, Security issues and measures: Protecting electronic commerce resources (Chapter 11), in *Electronic Commerce* (H. Bidgoli, Ed.). San Diego, CA: Academic Press, 2002, pp. 363-398.
- [7] J. Borges and M. Levene, Data mining of user navigation patterns, *Lecture Notes in Computer Science*, vol. 1836, pp 92-112, 2000.
- [8] P. K. Chan, W. Fan, A. L. Prodromidis, and S. J. Stolfo, Distributed data mining in credit card fraud detection, *IEEE Intelligent Systems*, vol. 14, no. 6, pp. 67-74, 1999.
- [9] S. Chaudhuri, U. Dayal and V. Narasayya, An overview of business intelligence technology, *Communications of the ACM*, vol 54, no. 8, pp. 88-98, 2011.
- [10] H. Chen and R. H. L. Chiang, Business Intelligence and analytics: From big data to big impact, *MIS Quarterly*, vol 36, no. 4, pp. 1165-1188, 2012.
- [11] Y.-L. Chong, Mobile commerce usage activities: The roles of demographic and motivation variables, *Technological Forecasting and Social Change*, vol. 80, no. 7, pp. 1350-1359, 2013.
- [12] E. Cohen, M. Datar, S. Fujiwara, A. Gionis, P. Indyk, R. Motwani, J. D. Ullman, and C. Yang, Finding interesting associations without support pruning, *IEEE Transactions Knowledge Data Engineering*, vol. 13, no. 1, pp. 64-78, 2001.
- [13] A. Doan, R. Ramakrishnan and A. Y. Halevy, Crowdsourcing systems on the world-wide web, *Communications of the ACM*, vol 54, no. 4, pp. 86-96, 2011.
- [14] European Central Bank, Report on Card Fraud, European Central Bank, Frankfurt am Main, 2012.
- [15] U. M. Fayyad, G. Piatetsky-Shapiro and P. Smyth, From data mining to knowledge discovery: An overview, in *Advances in Knowledge Discovery and Data Mining* (U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, Eds.). Menlo Park, CA: AAAI Press/MIT Press, 1996, pp. 1-34.
- [16] S. Goel, A. Broder, E. Gabrilovich, and B. Pang, Anatomy of the long tail, in *Proceedings of the Third ACM International Conference on Web Search and Data Mining - WSDM*, New York, 2010, pp. 201.
- [17] M. S. Hajmohammadi, R. Ibrahim and Z. A. Othman, Opinion mining and sentiment analysis: A survey, *International Journal of Computers & Technology*, vol. 2, no. 3, pp. 171-178, 2012.
- [18] J. Han, Data mining, in *Encyclopedia of Distributed Computing* (J. Urban and P. Dasgupta, Eds.). Norwell, MA: Kluwer Academic Publishers, 1999.
- [19] J. Han and Y. Fu, Mining multiple-level association rules in large databases, *IEEE Transactions on Knowledge and Data Engineering*, vol. 11, no. 5, pp. 798-805, 1999.
- [20] K. Jain, M. N. Murty and P. J. Flynn, Data clustering: a review, *ACM Computing Surveys*, vol. 31, no. 3, pp. 264-323, 1999.
- [21] A. Khan, T. Singh and A. Sinhal, Implement credit card fraudulent detection system using observation probabilistic in hidden markov model, in *Proceedings of Nirma University International Conference on Engineering (NUICONE)*, Ahmedabad, 2012, pp. 1-6.

- [22] K. Lai and N. Cerpa, Support vs. confidence in association rule algorithms, in Proceedings of the OPTIMA Conference, Curicó, 2001.
- [23] Y. Lu, H. He, Q. Peng, W. Meng, and C. Yu. Clustering e-commerce search engines based on their search interface pages using wise-cluster, *Data & Knowledge Engineering*, vol. 59, no. 2, pp. 231-246, 2006.
- [24] S.K. Madria, S.S. Bhowmick, W.K. Ng, E.P. Lim, Research issues in web data mining, *Lecture Notes in Computer Science*, vol. 1676, pp. 303-312, 1999.
- [25] R. Mikut and M. Reischl, Data mining tools, *WIRES: Data Min Knowl Discov*, vol. 1, no. 5, pp. 431-443, 2011.
- [26] M. Mohri, A. Rostamizadeh and A. Talwalkar, *Foundations of Machine Learning*. Cambridge, MA: The MIT Press, 2012.
- [27] R. Natarajan and B. Shekar, Interestingness of association rules in data mining: Issues relevant to e-commerce, *Sadhana*, vol. 30, no. 2-3, pp. 291-309, 2005.
- [28] E.W.T. Ngai, Y. Hu, Y.H. Wong, Y. Chen, and X. Sun, The application of data mining techniques in financial fraud detection: a classification framework and an academic review of literature, *Decision Support Systems*, vol. 50, no. 3, pp. 559-569, 2011.
- [29] K. Nusair and N. Hua, Comparative assessment of structural equation modeling and multiple regression research methodologies: E-commerce context, *Tourism Management*, vol. 31, no. 3, pp. 314-324, 2010.
- [30] B. Pang and L. Lee, Opinion mining and sentiment analysis, *Foundations and Trends in Information Retrieval*, vol. 2, no.1-2, pp. 1-135, 2008.
- [31] M.J. Pazzani, Knowledge discovery from data?, *IEEE Intelligent System*, vol. 15, no. 12, pp. 10-13, 2000.
- [32] C. Phua, V. Lee, K. Smith and R. Gayler. (2010, September) A comprehensive survey of data mining-based fraud detection research. *The Smithsonian/NASA Astrophysics Data System*. [Online]. Available: <http://adsabs.harvard.edu/abs/2010arXiv1009.6119P>
- [33] M.J. Shaw, C. Subramaniam, G.W. Tan, and M.E. Welge, Knowledge management and data mining for marketing, *Decision Support Systems*, vol. 31, no. 1, pp. 127-137, 2001.
- [34] D. Thorleuchter and D. V. den Poel. Predicting e-commerce company success by mining the text of its publicly-accessible website, *Expert Systems with Applications*, vol. 39, no. 17, pp. 13026-13034, 2012.
- [35] B. Thuraisingham, A Primer for understanding and applying data mining, *IT Professional*, vol. 2, no. 1, pp. 28-31, 2000.
- [36] M. Tsytarau, and T. Palpanas, Survey on mining subjective data on the web, *Data Mining and Knowledge Discovery*, vol. 24, no. 3, pp. 478-514, 2012.
- [37] R. Vedala and B. Kumar, An application of naive bayes classification for credit scoring in e-lending platform, in Proceedings International Conference on Data Science Engineering (ICDSE), Kochi, 2012, pp. 81-84.
- [38] T. Verbraken, F. Goethals, W. Verbeke, and B. Baesens. Using social network classifiers for predicting e-commerce adoption, in *E-Life: Web-Enabled Convergence of Commerce, Work, and Social Life* (M. Shaw, D. Zhang, and W. Yue, Eds.). Springer Berlin Heidelberg, 2012, pp. 9-21.
- [39] L. Wang and H. Xu. The application research of crm in e-commerce based on association rule mining. In *Advances in Future Computer and Control Systems* (D. Jin and S. Lin, Eds.). Springer Berlin Heidelberg, 2012, pp. 51-56.
- [40] J. Wu, Q. Liu and S. Luo, Clustering technology application in e-commerce recommendation system, in Proceedings ICMECG International Conference Management of e-Commerce and e-Government, Jiangxi, 2008, pp. 200-203.
- [41] X. Wu, V. Kumar, J. Ross Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A. Ng, B. Liu, P. S. Yu, Z.-H. Zhou, M. Steinbach, D. J. Hand, and D. Steinberg. Top 10 algorithms in data mining, *Knowl. Inf. Syst.*, vol. 14, no. 1, pp. 1-37, 2007.
- [42] H. Xiong, P.-N. Tan and V. Kumar, Hyperclique pattern discovery, *Data Min. Knowl. Discov.*, vol. 13, no. 2, pp. 219-242, 2006.
- [43] Q. Yang and X. Wu, 10 challenging problems in data mining research, *International Journal Information Technology & Decision Making*, vol. 5, no. 4, pp. 597-604, 2006.
- [44] X.-Z. Zhang, Building personalized recommendation system in e-commerce using association rule-based mining and classification, in Proceedings International Conference on Machine Learning and Cybernetics, Hong Kong, 2007, pp. 4113-4118.
- [45] X. Zheng, S. Zhu and Z. Lin, Capturing the essence of word-of-mouth for social commerce: Assessing the quality of online e-commerce reviews by a semi-supervised approach, *Decision Support Systems*, vol. 56, no. 0, pp. 211-222, 2013.
- [46] X. Zhu, Semi-supervised learning literature survey, *Computer Sciences*, University of Wisconsin-Madison, Technical Report 1530, 2005.