Discovering interesting rules from financial data

Przemysław Sołdacki
Institute of Computer Science
Warsaw University of Technology
Ul. Andersa 13, 00-159 Warszawa
Tel: +48 609129896
email: psoldack@ii.pw.edu.pl

Abstract

In this paper problem of mining data with weights and finding association rules is presented. Some applications are discussed, especially focused on financial data. Solutions of the problem are analyzed. A few approaches are proposed and compared. Pruning based on measures of rules interestingness is described and some measures proposed in literature are shown. Influence of data weights on these measures is also discussed.

1. Introduction

Discovering association rules is one of the most important tasks in data mining and many efficient algorithms were proposed in literature. However, the number of discovered rules is often so large, so the user cannot analyze all discovered rules. To overcome that problem several methods for mining interesting rules only have been proposed. One of them is pruning based on interestingness measures. Many measures have been proposed in literature. We describe and compare them.

Most of data mining algorithms assume equal data weights of all transactions. It is reasonable in most cases. However sometimes different data weights can increase applicability and accuracy of algorithms. Data weights influence on interestingness measures. In this paper we describe how weights of transaction items can be used for mining interesting rules, especially in financial data.

1.1. Layout

The rest of this paper is organized in the following manner. Section 2 summarizes related work. Section 3 formally defines association rules and their properties. Section 4 describes usage of weight of items in discovering frequent itemsets and association rules. Section 5 presents association rules interestingness measures proposed in literature. Section 6 concludes this paper.

2. Related Work

The efficient algorithms for finding all association rules were proposed in [2,9]. In [4,6] the problem of constraint-based mining in dense data was investigated. In [6] an algorithm for mining all association rules with given consequent meeting specified by the user conditions on minimal support, confidence and improvement was introduced. The *improvement* of a rule was defined as the minimum difference between its confidence and the confidence of any proper sub-rule with the same consequent.

For discovering the most interesting rules varied metrics including coverage, lift, conviction or PS (Piatetsky-Shapiro measure) were used. A new approach to the problem of finding the optimal rules, which involves a partial order on rules defined in terms of both rule support and confidence, was defined in [5]. That concepts of rule interestingness capture the best rules according to previously mentioned measures. Many other interestingness measures have been proposed in literature. Some of them can be found in [6],[13],[14],[15],[16]. Some approaches to using data weights can also be found in [12].

Another approach facilitating mining of interesting association rules is based on finding such a subset of all rules, which enables inferring all of them. In [7] a *cover operator* of rule was introduced. A *Cover* of rule $X \Rightarrow Y$, $X \neq \emptyset$, $Y \neq \emptyset$ is defined as following: $Cover(X \Rightarrow Y) = \{X \cup Z \Rightarrow V \mid Z,V \subseteq Y \text{ and } X \cap Y = \emptyset \text{ and } V \neq \emptyset\}$. By means of cover the set of representative rules can be defined as set of rules where each rule *r* does not belong to cover of any other rule. The efficient algorithm for finding representative rules was presented in [8]. A concept of closed frequent itemsets and a method for generating non-redundant rules based on that concept was described in [10]. A rule *r* is redundant if there exists a rule with the same support and confidence as *r* and either its consequent is a subset of consequent of rule *r* or its antecedent is a subset of antecedent of rule *r*.

3. Association rules and their properties

We begin with definition of necessary terminology. A database *D* is a set of transactions, which are sets over a finite item domain *I*. Let *k-itemset* be a set of *k* items from database.

The most basic property of an itemset is *support*. It is defined as percentage of transactions in *D* database, which contain given itemset. It is referred as a relative support. Formal expression is shown below:

$$support(A) = |\{T \in D \mid A \subseteq T\}| \,/\, |D|, \text{ where:}$$

A – itemset, *T* – transaction, *D* - database

Sometimes an absolute support is used. It is defined as: $support_a(A) = |\{T \in D \mid A \subseteq T\}|$

Frequent itemset is an itemset with support not less than a given minimal level called *minSup*. An itemset is maximal frequent if it have no frequent superset.

Association rule is an implication:

$$X \Rightarrow Y, \text{ where } X, Y \text{ are itemsets over } I \text{ and } X \neq \emptyset,\ Y \neq \emptyset \text{ and } X \cap Y = \emptyset.$$

X is called an antecedent of rule, Y is called consequent of rule. The support of rule $X \Rightarrow Y$ is equal to the

$support(X \cup Y)$. Confidence of rule $X \Rightarrow Y$ denoted as $confidence(X \Rightarrow Y)$ is defined as:

$$confidence(X \Rightarrow Y) = support(X \Rightarrow Y) \,/\, support(X).$$

Parameter *minConf* is defined by the users and indicates minimal confidence that discovered rules need to have.

Support properties

Let A and B, be itemsets over D database. Then the following property is kept:

$$A \subseteq B \;\Rightarrow\; support(A) \geq support(B)$$

It is implied directly by support definition. A number of transactions containing an itemset is less or equal to a number of transactions containing its subset.

This property implies that every subset of a frequent itemset is also frequent. It is very important and useful fact for frequent itemset discovery and is utilized by almost all algorithms.

4. Data with weights

4.1. Applications

One of classic knowledge exploration problems is purchase-basket analysis. Let us discover it from salesman point of view where income or profit is the key issue. Number of transactions is not so important. It is assumed that transactions giving higher income are more interesting than others. Consequently we presume that a day of highest income is more important than a day of maximal number of transactions.

Usually when analyzing purchase-basket we can easily get goods along with their prices and sometimes even with margin on every product. We can use all these information. We can use prices as data weights if we are more interested in income. Using margin we will set store by profit.

Certainly there are many other applications, where data weighting is helpful. It includes almost all situations, when data being explored are associated with money.

Also other values and measures can be used as data weights, for instance: a time of activities, a size or physical weight of goods and so on.

4.2. Data formats

There are two main approaches of data weighting depending on data formats:

- associating weights to transaction items
- associating weights to transactions

Let us present following examples in different data formats:

Table 1: Data A. Relational format, weights associated to transactions

Transaction ID (TID)	X1	L	Weight
1	A	4	8
2	B	3	9
3	A	3	2
4	C	1	10
5	A	8	11

Table 2: Data B. Transactional format, weights associated to items

Transaction ID (TID)	Item	Weight
1	A	5
1	B	3
2	B	3
2	D	6
3	E	1
3	D	1

As we can see a data format naturally selects resolution of information about weights. Certainly there are also other possibilities. We can imagine some hybrid formats. Next two tables show such examples.

Table 3: Data C. Relational format, weights associated to items (weights after colon)

Transaction ID (TID)	X1	L
1	A:5	4:3
2	B:3	3:6
3	A:1	3:1
4	C:6	1:4
5	A:6	8:5

Table 4: Data D: Transactional format, weights associated to transaction (value of WEIGHT attribute)

Transaction ID (TID)	Item
1	X1=A
1	L=4
1	WEIGHT =3
2	X1=B
2	L=3
2	WEIGHT =6
3	X1=A
3	L=3
3	WEIGHT =4

Let weights be combined by arithmetic adding. It is true in many kinds of weights, for example money. To have a weight of transaction we just need to add weights of all items contained in it. Certainly it is not easy and usually not even possible to get items weights from transaction weights. In certain cases analytical methods can be used.

4.3. Itemsets and association rules with weights

Support is a relevance measure of an itemset. By modifying definition of support we can use information on data weights. Certainly it influences support and confidence of association rules, which are based on itemsets support.

4.3.1. Item weights

Assume following support definition:

$support(A) = \Sigma_{t\in D}\, \Sigma_{\{e \,|\, e \in\, t \,\cap A\}}\, weight(e)$,

where D – database, t – transaction, e – item

Such a definition represents sum of weights of all items contained in a set. Let us consider purchase data, for instance. In this situation the support shows total amount of money earned by selling goods contained in an itemset. Unfortunately such a definition is very inconvenient, because increasing cardinality of an itemset increases support. It violates one of main support's properties. We would like to have a definition where increasing cardinality of an itemset effects in equal or lower value of support. This property is very important and is used by almost all algorithms for finding frequent itemsets and association rules. Therefore we have to consider changes in this definition to achieve needed property.

4.3.2. Transaction weights

Let us define itemset support as:

$$support(A) = \Sigma_{\{t \mid t \in D \wedge A \subseteq t\}} \Sigma_{e \in t}\, weight(e)$$

what is equivalent to

$$support(A) = \Sigma_{\{t \mid t \in D \wedge A \subseteq t\}}\, weight(t)$$

where $weight(t) = \Sigma_{e \in t}\, weight(e)$.

It means that support if itemset is a sum of weights of transactions containing given itemset. In purchase-basket example support represent summary sale of transactions that include items of given set. It is possible that a very cheap item is frequently present in very expensive transactions. However such information is also valuable, because maybe this cheap item increases sale of other expensive items, which appear in a same transaction. It can be a good reason to make a promotion and to give such cheap item for free.

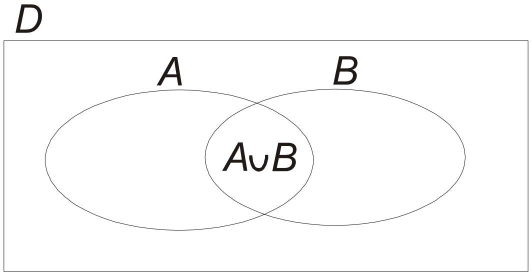
An additional advantage of such a definition is that assuming all transaction weights equal to 1, we get classic support definition.

5. Interestingness measures

Support and confidence are the most basic measures of rules interestingness. Usually interesting rules are defined as rules describing surprising uncommon situations. In such cases support and confidence are not sufficient. Many additional measures were proposed.

Please note that confidence definition is based on support of entire rule and of antecedent. It does not use support of a consequent. Thus we lost some information. It can cause some negative effects. For instance: if customers buy milk in 80% of transactions and it is an independent event on buying salmon then confidence of *salmon* ⇒ *milk* rule is 80%. Certainly rule describing independent events is not interesting.

Most of measures are defined by combination of itemsets probabilities. Support of itemset describes probability of its appearance in transaction from database *D*. Itemsets of rule $A \Rightarrow B$ are shown below:



We can define probabilities of itemsets appearances in the following manner:

$P(A) = support(A)$
$P(B) = support(B)$
$P(A,B) = support(A \cup B)$
$P(B \mid A) = P(A,B) / P(A) = confidence(A \Rightarrow B)$
$P(\sim A) = 1\text{-}P(A) = 1\text{-}support(A)$, where ~ is a negation

Coverage [13]

$$coverage(A \Rightarrow B) = P(A,B) / P(B) = support(A \cup B) / support(B)$$

It shows what part of itemsets from consequent is covered by a rule. Its values are in range [0; 1].

Lift [6],[14]

This measure is also called *interest,* defined as follows:

$$lift(A \Rightarrow B) = P(A,B) / (P(A) \cdot P(B)) = support(A \cup B) / (support(A) \cdot support(B))$$

It is equal to proportion of real support of itemset $A \cup B$ to expected support (assuming independent events). Therefore it shows level of correlation between antecedent and consequent. However it does not let to determine direction of implication, because it is symmetric.

Lift can be also defined using *confidence*:

$$lift(A \Rightarrow B) = confidence(A \Rightarrow B) / support(B)$$

This form of definition leads to another interpretation. *Lift* shows proportion of conditional probability *B* (under condition of *A*) to unconditional probability of *B*. It is explained in the following example:

Let us presume that during exploration process rule *bread* ⇒ *milk* is found and its confidence is equal to 80%. It potentially holds useful knowledge. However, if 90% of all customers buy milk then this rule is not interesting. In such a case *lift* is a very helpful measure. In considered example its value is lower than 1.

If antecedent and consequent were independent then:

$$confidence(A \Rightarrow B) = support(A \cup B) / support(A) =$$
$$= (support(A) \cdot support(B)) / support(A) = support(B)$$

Thus expected confidence is equal to support of consequent. Then *lift* shows how unexpected is real *confidence*.

Its values are in range [0;+∞). Values lower than 1 mean, that satisfying condition of antecedent decreases probability of consequent in comparison to unconditional probability. Consequently, values higher than 1 mean, that satisfying condition of antecedent increases probability of consequent in comparison to unconditional probability. If antecedent and consequent are independent then *lift* is equal to 1.

Piatetsky-Shapiro [13]

$$PS(A \Rightarrow B) = P(A,B) - P(A) \cdot P(B) = support(A \cup B) - support(A) \cdot support(B)$$

Absolute value of this measure shows dependence between antecedent and consequent. Its values are in range <-1;+1>. Positive values mean that conditional probability of consequent (under condition from antecedent) is higher than unconditional one. If antecedent and consequent are independent *PS* measure is equal to 0.

Conviction [14]

$$conviction(A \Rightarrow B) = P(A)\ P(\sim B) / P(A,\sim B)$$

This measure was derived from implication definition. Implication $A \Rightarrow B$, can be presented in following form $\sim(A \wedge \sim B)$. This form was transformed by avoiding negation and whole term was transferred to denominator. Thus the measure shows level of dependence between *A* and *~B*. After some transformations we achieve:

$$conviction(A \Rightarrow B) = P(A)\ (1\text{-}P(B)) / (P(A)\text{-}P(A,B))$$

Using supports instead of probabilities leads to formula shown below:

$$conviction(A \Rightarrow B) = support(A) \cdot (1\text{-}\ support(B)) / (support(A)\text{-}\ support(A \cup B))$$

what is equivalent to:

$$conviction(A \Rightarrow B) = (1\text{-}\ support(B)) / (1\text{-}confidence(A \Rightarrow B))$$

Its values are in range [0;+∞]. If antecedent and consequent are independent it is equal to 1. For implications occurring in all cases measure's value is equal to +∞.

Interestingness [6]

This measure is extension of *lift* measure. It shows level of interestingness based on support of antecedent and consequent, correlations between them and two additional parameters. It is defined as:

$$I(A \Rightarrow B) = \left(\left(\frac{\sup(A \cup B)}{\sup(A) \cdot \sup(B)} \right)^{k} - 1 \right) \cdot \left(\sup(A) \cdot \sup(B) \right)^{m}$$

where parameter k – importance of event dependency, m – importance of support.

J-measure [16]

This measure shows how much information is contained in a rule. This measure combining it with consequent support shows how much rule is interesting.

5.1. Interestingness measures with weights

Using support definition from section 4.3.2 we can calculate other measures of association rule interestingness. Let us consider measure values with and without weights for some data shown below. Transaction 3 is weighted 11, which is sum of item weights. Please notice that without weights these data are symmetric considering items A and B, but with weights they are not.

Table 5: Sample data

Transaction ID (TID)	Item	Weight
1	A	1
2	B	10
3	A	1
3	B	10
4	A	1
5	B	10

Table 6: Interestingness measures

Measure	Value without weights	Value with weights
Support(A)	60%	39%
Support(B)	60%	94%
Support(A∪B)	20%	33%
Confidence(A⇒B)	33%	85%
Confidence(B⇒A)	33%	35%
Coverage(A⇒B)	60%	35%
Coverage(B⇒A)	60%	85%
Lift(A⇒B)	0.56	0.9
Lift(B⇒A)	0.56	0.9
PS(A⇒B)	-16%	-3.66%
PS(B⇒A)	-16%	-3.66%
Conviction(A⇒B)	0.6	0.39
Conviction(B⇒A)	0.6	0.94

Very interesting are asymmetric measures: confidence and conviction. Confidence is higher for rules with implication from item with lower weight to item with higher weight. In many cases it is very reasonable. During purchase basket data mining implication from a cheap product to expensive one is more interesting than the opposite implication.

Unexpectedly conviction measure behaves the opposite way. Certainly it sometimes also can be useful, however it is important to keep it in mind.

6. Conclusions

We discussed some aspects of mining interesting association rules. We analyzed two approaches to processing data with weights and presented advantages and disadvantages of both. Basic and additional measures of rule interestingness was collected and presented. We also highlighted weights influence to interestingness measures. Also several applications were proposed especially focused on financial aspects.

Using weights is a very interesting approach in association rules discovery process. In this paper only financial data analysis was discussed. It seems that there are many other important applications and it is possible direction of future extensions.

References

[1] Agrawal R., Imielinski T., Swami A.: Mining Associations Rules between Sets of Items in Large Databases; Proc. of the ACM SIGMOD Conf. on Management of Data. Washington DC 1993.

[2] Agrawal R., Srikant R.: Fast Algorithms for Mining Association Rules in Large Databases Int'l Conf. on VLDs; Santiago, Chile, 1994.

[3] Agrawal R., Srikant R.: Mining Association Rules. Proc. of the 21th VLDB Conf., Switzerland 1995

[4] Bayardo R.J. Jr.: Efficiently Mining Long Patterns from Databases; ACM-SIGMOD Int'l Conf. on Management of Data 1998.

[5] Bayardo R.J. Jr., Agrawal R.: Mining the Most Interesting Rules; ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining1999.

[6] Bayardo R.J. Jr., Agrawal R., Gunopulos D.: Constraint-Based Mining in Large, Dense Databases; Int'l Conf. on Data Engineering, 1999.

[7] Kryszkiewicz M.: Representative Association Rules; Proc. of PAKDD 1998 Melbourne, Australia

[8] Kryszkiewicz M.: Representative Association Rules and Minimum Condition Maximum Consequence Association Rules; Proc. of PAKDD 1998, France

[9] Savasere A., Omiecinski E., Navathe S.: An Efficient Algorithm for Mining Association Rules in Large Databases; In Proc. of the 21st Int'l Conf. on Very Large Data-Bases Zurich 1995.

[10] Toivonen H., Klemettinen M., Ronkainen P., Hätönen K., Mannila H.: Pruning and Grouping Discovered Association Rules; Mlnet Workshop on Statistics, Machine Learning and Discovery in Databases; Heraklion, Create, Greece 1995

[11] Zaki M.J.: Generating Non-Redundant Association Rules; KDD 2000 Boston MA USA

[12] Hilderman, R. J., & Hamilton, H. J. (2001). Evaluation of interestingness measures for ranking discovered knowledge. Proc. of the 5th Pacific-Asia Conference on Knowledge Discovery and Data Mining (pp. 247--259). China: Springer-Verlag.

[13] B. Iglesia "Induction of Interesting Rules from Large Datasets", University of East Anglia, 1999.

[14] S. Brin, R. Motwani, J. D. Ullman, S. Tsur „Dynamic Itemset Counting and Implication Rules for Market Basket Data", 1997.

[15] B. Gray, M.E. Orlowska "Clustering categorical attributes into interesting association rules". Proc. of the 2nd Pacific-Asia Conference on PAKDD'1998.

[16] P. Smyth and R.M. Goodman "Rule induction using information theory". In Knowledge Discovery in Databases, 1991.