In 5th International Conference on Data Mining, Text Mining and their Business Applications, 2004

Association rules model of e-banking services

Vasilis Aggelis
Department of Computer Engineering and Informatics, University of Patras, Greece

Abstract

The introduction of data mining methods in the banking area although conducted in a slower way than in other fields, mainly due to the nature and sensitivity of bank data, can already be considered of great assistance to banks as to prediction, forecasting and decision making. One particular method is the investigation for association rules between products and services a bank offers. Results are generally impressive since in many cases strong relations are established, which are not easily observed at a first glance. These rules are used as additional tools aiming at the continuous improvement of bank services and products helping also in approaching new customers. Scope of this paper is the presentation of a model which discovers and determinate such rules concerning e-banking using different methods.
Keywords: data mining, e-banking, association rules, a priori, generalized rule induction

1 Introduction

Determination of association rules concerning bank data is a challenging though demanding task since:

- The volume of bank data is enormous. Therefore the data should be adequately prepared by the data miner before the final step of the application of the method.
- The objective must be clearly set from the beginning. In many cases not having a clear aim results in erroneous or no results at all.
- Good knowledge of the data is a prerequisite not only for the data miner but also for the final analyst (manager). Otherwise wrong results will be produced by the data miner and unreliable conclusions will be drawn by the manager.

- Not all rules are of interest. The analyst should recognize powerful rules as to decision making.

The challenge of the whole process rises from the fact that the relations established are not easily observed without use of data mining methods.

The increase of electronic transactions during the last years is quite rapid. E-banking nowadays offers a complete sum of products and services facilitating not only the individual customers but also corporate customers to conduct their transactions easily and securely. The ability to discover rules between different electronic services is therefore of great significance to a bank.

Identification of such rules offers advantages as the following:

- Description and establishment of the relationships between different types of electronic transactions.
- The electronic services become more easily familiar to the public since specific groups of customers are approached, those use specific payment manners.
- Customer approach is well designed with higher possibility of successful engagement.
- The improvement of already offered bank services is classified as to those used more frequently.
- Reconsidering of the usefulness of products exhibiting little or no contribution to the rules.

Of certain interest is the certification of the obtained rules using other methods of data mining to assure their accuracy.

This study is concerned with the identification of rules between several different ways of payment. The software used is SPSS Clementine 7.0. Section 2 contains association rules' basic features and algorithms, while in section 3 the process of investigation for association rules is described. Experimental results can be found in section 4 and finally section 5 contains the main conclusions of this work accompanied with future work suggestions in this area.

2 Association rules basics

A rule consists of a left-hand side proposition (antecedent) and a right-hand side (consequent) [3]. Both sides consist of boolean statements. The rule states that if the left-hand side is true, then the right-hand is also true. A probabilistic rule modifies this definition so that the right-hand side is true with probability p, given that the left-hand side is true.

A formal definition of association rule [6, 7, 14, 17, 18, 20] is given below.

Definition. An association rule is a rule in the form of

$$X \rightarrow Y \quad (1)$$

Where *X* and *Y* are predicates or set of items.

As the number of produced associations might be huge, and not all the discovered associations are meaningful, two probability measures, called support and confidence, are introduced to discard the less frequent associations in the database. The support is the joint probability to find X and Y in the same group; the confidence is the conditional probability to find in a group Y having found X.

Formal definitions of support and confidence [1, 3, 14, 18, 20] are given below.

Definition Given an itemset pattern X, its frequency fr(X) is the number of cases in the data that satisfy X. Support is the frequency $fr(X \wedge Y)$. Confidence is the fraction of rows that satisfy Y among those rows that satisfy X,

$$c(X \rightarrow Y) = \frac{fr(X \wedge Y)}{fr(X)} \tag{2}$$

In terms of conditional probability notation, the empirical accuracy of an association rule can be viewed as a maximum likelihood (frequency-based) estimate of the conditional probability that Y is true, given that X is true.

2.1 A priori algorithm

Association rules are among the most popular representations for local patterns in data mining. A priori algorithm [1, 2, 3, 6, 10] is one of the earliest for finding association rules. This algorithm is an influential algorithm for mining frequent itemsets for boolean association rules. This algorithm contains a number of passes over the database. During pass k, the algorithm finds the set of frequent itemsets L_k of length k that satisfy the minimum support requirement. The algorithm terminates when L_k is empty. A pruning step eliminates any candidate, which has a smaller subset. The pseudo code for a priori algorithm is following:

```
Ck: candidate itemset of size k
Lk: frequent itemset of size k
L1 = {frequent items};
For (k=1; Lk != null; k++) do begin
        Ck+1 = candidates generated from Lk;
        For each transaction t in database do
                Increment the count of all candidates in
                Ck+1 that are contained in t
        Lk+1 = candidates in Ck+1with min_support
        End
Return Lk;
```

2.2 Generalized rule induction (GRI)

In the generalized rule induction task [3, 4, 5, 11, 14, 15, 16], the goal is to discover rules that predict the value of a goal attribute, given the values of other attributes. However the goal attribute can be any attribute not occurring in the antecedent of the rule.

3 Investigation for association rules

Scope of this study is the determination of relations between different payment types e-banking users conduct. These relations are of great importance to the bank as they enhance the selling of its products, by making the bank more aware of its customer's behavior and the payment way they prefer. They also contribute in improvement of its services.

Namely the following types of payments were used:

1. DIASTRANSFER payment orders – Funds transfers between banks that are participating in DIASTRANSFER system (Greek funds transfer system).
2. SWIFT payment orders – Funds transfers among national and foreign banks.
3. Funds transfers – Funds transfers inside the bank.
4. Forward funds transfers – Funds transfers inside the bank with forward value date.
5. Greek Telecommunications Organization (GTO) and Public Power Corporation (PPC) standing orders – Standing orders for paying GTO and PPC bills.
6. Social Insurance Institute (SII) payment orders – Payment orders for paying employers' contributions.
7. VAT payment orders.
8. Credit card payment orders.

All above payments concerning Greek e-banking market but the majority of them has a similarity with other countries e-banking markets.

For the specific case of this study a random sample of e-banking customers, both individuals and companies, was used. A boolean value is assigned to each different payment type depending on whether the payment has been conducted by the users or not. In case the user is individual field "company" receives the value 0, else 1.

A sample of the above data set is shown in table 1. We use the following names for the fields: Co as company, D as DIASTRANSFER payment orders, S as SWIFT payment orders, F as funds transfer, FF as forward funds transfer, GP as GTO and PPC standing orders, SII as SII payment orders, VAT as VAT payment orders and CC as credit card payment orders

In order to discover association rules the a priori [4] and generalized rule induction (GRI) [4] methods were used. These rules are statements in the form if antecedent then consequent.

Table 1: Data Set Sample

User	Co	D	S	F	FF	GP	SII	VAT	CC
…	…	…	…	…	…	…	…	…	…
User103	0	T	F	T	T	F	F	F	T
User104	1	T	T	T	T	F	T	T	F
User105	1	T	F	T	T	T	T	T	T
User106	0	T	F	T	F	T	F	F	T
…	…	…	…	…	…	…	…	…	…

4 Experimental results

By the use of a priori method and setting minimum rule support = 10 and minimum rule confidence = 20, eleven rules were determined and shown in fig. 1.

Using method GRI with minimum rule support = 10 and minimum rule confidence = 40 resulted in the four rules seen in fig. 2.

After comparison between the rules obtained by the two methods it can be concluded that the most powerful is:

- If SII payment order then VAT payment order (confidence=81). (Rule 1)

Additionally it can be observed that this rule is valid for the reverse case too, namely:

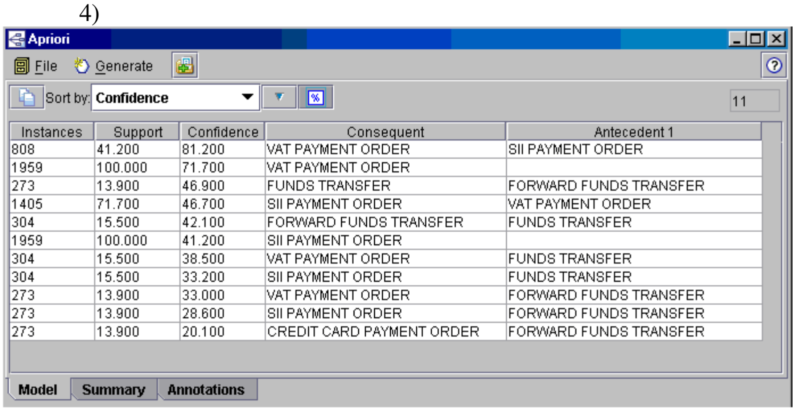
- If VAT payment order then SII payment order (confidence= 47). (Rule 2)

Two other rules obtained exhibiting confidence > 40 are the following:

- If funds transfers then forward funds transfers (confidence=47), (Rule 3)

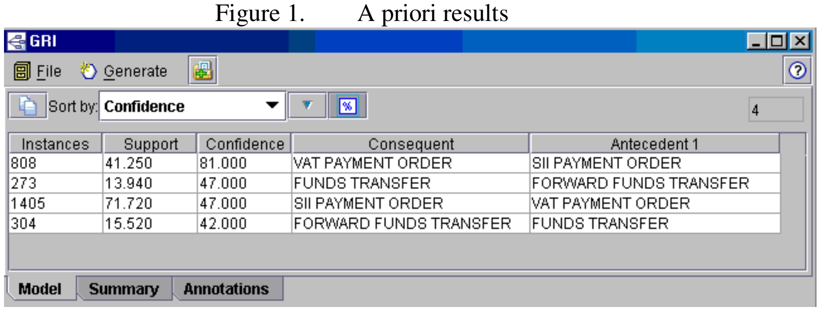
valid also reversely.

- If forward funds transfers then funds transfers (confidence=42), (Rule 4)



Instances	Support	Confidence	Consequent	Antecedent 1
808	41.200	81.200	VAT PAYMENT ORDER	SII PAYMENT ORDER
1959	100.000	71.700	VAT PAYMENT ORDER	
273	13.900	46.900	FUNDS TRANSFER	FORWARD FUNDS TRANSFER
1405	71.700	46.700	SII PAYMENT ORDER	VAT PAYMENT ORDER
304	15.500	42.100	FORWARD FUNDS TRANSFER	FUNDS TRANSFER
1959	100.000	41.200	SII PAYMENT ORDER	
304	15.500	38.500	VAT PAYMENT ORDER	FUNDS TRANSFER
304	15.500	33.200	SII PAYMENT ORDER	FUNDS TRANSFER
273	13.900	33.000	VAT PAYMENT ORDER	FORWARD FUNDS TRANSFER
273	13.900	28.600	SII PAYMENT ORDER	FORWARD FUNDS TRANSFER
273	13.900	20.100	CREDIT CARD PAYMENT ORDER	FORWARD FUNDS TRANSFER

Figure 1. A priori results



Instances	Support	Confidence	Consequent	Antecedent 1
808	41.250	81.000	VAT PAYMENT ORDER	SII PAYMENT ORDER
273	13.940	47.000	FUNDS TRANSFER	FORWARD FUNDS TRANSFER
1405	71.720	47.000	SII PAYMENT ORDER	VAT PAYMENT ORDER
304	15.520	42.000	FORWARD FUNDS TRANSFER	FUNDS TRANSFER

Figure 2. GRI results

As seen, there exists strong relationship between VAT payment order and SII payment order, as is the case between funds transfers and forward funds transfers.

These strong relationships are visualized on the web graph of fig. 3 and the tables of its links in fig. 4.

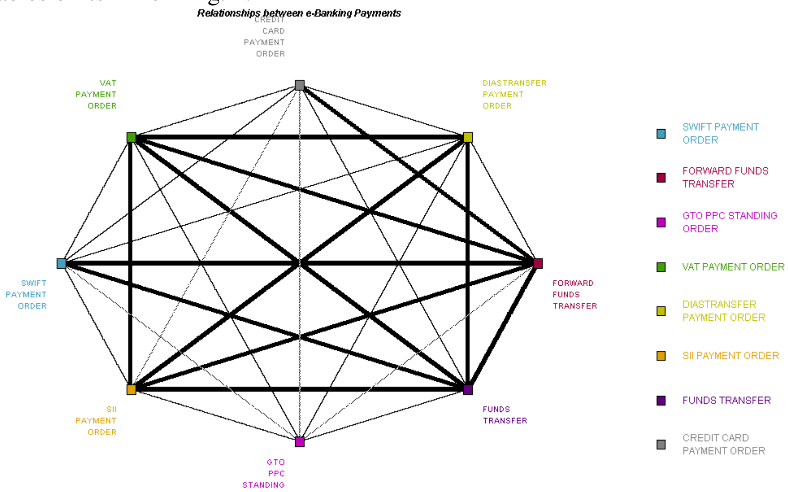


Figure 3. Web graph

Strong Links

Links	Field 1	Field 2
656	VAT PAYMENT ORDER = "T"	SII PAYMENT ORDER = "T"
128	FUNDS TRANSFER = "T"	FORWARD FUNDS TRANSFER = "T"
117	FUNDS TRANSFER = "T"	VAT PAYMENT ORDER = "T"
101	FUNDS TRANSFER = "T"	SII PAYMENT ORDER = "T"
90	FORWARD FUNDS TRANSFER = "T"	VAT PAYMENT ORDER = "T"
78	FORWARD FUNDS TRANSFER = "T"	SII PAYMENT ORDER = "T"
55	FORWARD FUNDS TRANSFER = "T"	CREDIT CARD PAYMENT ORDER = "T"
47	SWIFT PAYMENT ORDER = "T"	FORWARD FUNDS TRANSFER = "T"
40	DIASTRANSFER PAYMENT ORDER = "T"	SII PAYMENT ORDER = "T"
37	DIASTRANSFER PAYMENT ORDER = "T"	FUNDS TRANSFER = "T"
37	SWIFT PAYMENT ORDER = "T"	FUNDS TRANSFER = "T"
36	DIASTRANSFER PAYMENT ORDER = "T"	VAT PAYMENT ORDER = "T"

Medium Links

Links	Field 1	Field 2
33	SWIFT PAYMENT ORDER = "T"	VAT PAYMENT ORDER = "T"
31	SWIFT PAYMENT ORDER = "T"	SII PAYMENT ORDER = "T"
27	FUNDS TRANSFER = "T"	CREDIT CARD PAYMENT ORDER = "T"
26	CREDIT CARD PAYMENT ORDER = "T"	VAT PAYMENT ORDER = "T"
25	DIASTRANSFER PAYMENT ORDER = "T"	FORWARD FUNDS TRANSFER = "T"
25	DIASTRANSFER PAYMENT ORDER = "T"	SWIFT PAYMENT ORDER = "T"
24	VAT PAYMENT ORDER = "T"	GTO PPC STANDING ORDER = "T"
17	SII PAYMENT ORDER = "T"	GTO PPC STANDING ORDER = "T"
17	FUNDS TRANSFER = "T"	GTO PPC STANDING ORDER = "T"
17	SWIFT PAYMENT ORDER = "T"	CREDIT CARD PAYMENT ORDER = "T"
17	DIASTRANSFER PAYMENT ORDER = "T"	CREDIT CARD PAYMENT ORDER = "T"
15	DIASTRANSFER PAYMENT ORDER = "T"	GTO PPC STANDING ORDER = "T"

Weak Links

Links	Field 1	Field 2
13	CREDIT CARD PAYMENT ORDER = "T"	SII PAYMENT ORDER = "T"
10	CREDIT CARD PAYMENT ORDER = "T"	GTO PPC STANDING ORDER = "T"
7	FORWARD FUNDS TRANSFER = "T"	GTO PPC STANDING ORDER = "T"
6	SWIFT PAYMENT ORDER = "T"	GTO PPC STANDING ORDER = "T"

Figure 4. Web graph links

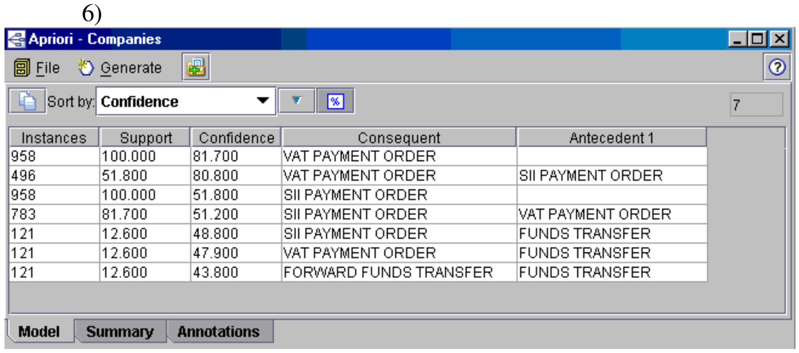
Interestingness [8, 9, 12, 13, 18, 19] of the above rules is high bearing in mind that generally a pattern or rule can be characterized interesting if it is easily understood, unexpected, potentially useful and actionable or it validates some hypothesis that a user seeks to confirm. Given the fact that payment types used in the rules exhibiting the highest confidence (VAT payment order, SII payment order) are usually payments conducted by companies and freelancers, a smaller subdivision of the whole sample was considered containing only companies for which the methods a priori και GRI were applied. The results are described below.

Using a priori method and setting minimum rule support = 10 and minimum rule confidence = 40, the seven rules shown in fig. 5 were obtained.

GRI method with minimum rule support = 10 and minimum rule confidence = 50 resulted in two rules, see fig. 6.

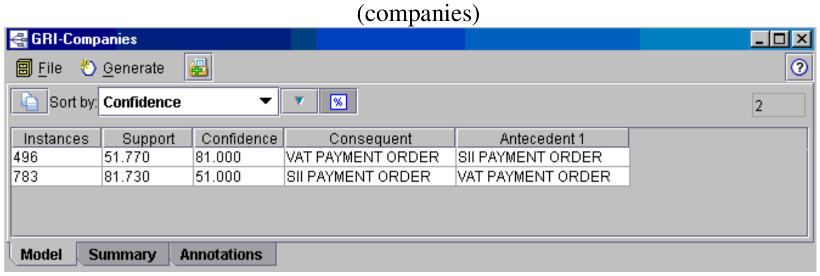
The comparison of the above rules leads to the conclusion that the validity of rules 1 and 2 is confirmed exhibiting even increased support:

- If SII payment order then VAT payment order (confidence=81). (Rule 5)
- If VAT payment order then SII payment order (confidence= 51). (Rule 6)



Instances	Support	Confidence	Consequent	Antecedent 1
958	100.000	81.700	VAT PAYMENT ORDER	
496	51.800	80.800	VAT PAYMENT ORDER	SII PAYMENT ORDER
958	100.000	51.800	SII PAYMENT ORDER	
783	81.700	51.200	SII PAYMENT ORDER	VAT PAYMENT ORDER
121	12.600	48.800	SII PAYMENT ORDER	FUNDS TRANSFER
121	12.600	47.900	VAT PAYMENT ORDER	FUNDS TRANSFER
121	12.600	43.800	FORWARD FUNDS TRANSFER	FUNDS TRANSFER

Figure 5. A priori results (companies)



Instances	Support	Confidence	Consequent	Antecedent 1
496	51.770	81.000	VAT PAYMENT ORDER	SII PAYMENT ORDER
783	81.730	51.000	SII PAYMENT ORDER	VAT PAYMENT ORDER

Figure 6. GRI results (companies)

5 Conclusions and Future work

In the present paper the way of discovery of association rules between different types of e-banking payment offered by a bank is described along with experimental results.

The basic outcome is that VAT payment orders and SII payment orders are the most popular and interconnected strongly. The detection of such relationships offers a bank a detailed analysis that can be used as a reference point for the conservation of the customer volume initially but also for approaching new customer groups who conduct these payments in order to increase its customer number and profit.

Our work was adopted by a Greek bank, which has extended the customer base in groups related to VAT and SII transactions, as companies, freelancers and accountants, resulting in the sizes increase that it is presented in table 2.

These modifications improved bank function in two ways:

Firstly, profit increases through the commissions from organizations and manipulation of the capitals for a short period of time (valeur days), and secondly functional cost reduces, since internet transactions cost is the lowest one.

Our models gradually become part of the bank function, while they can be adopted by other banks and organisations in Greece and abroad.

Similar patterns can be derived after analysis of payment types of specific customer groups resulting from various criteria either gender, residence area, age, profession or any other criteria the bank assumes significant. The rules to be discovered show clearly the tendencies created in the various customer groups helping the bank to approach these groups as well as re-design the offered electronic services in order to become more competitive.

In addition to the two methods used, detection of association rules can employ several other data mining methods such as decision trees in order to confirm the validity of the results.

Table 2: Increasing sizes

Payment Type	+/- (2003 vs 2002)
VAT Payment (Num. of)	19.00%
VAT Payment (Amount)	19.00%
SII Payment (Num. Of)	75.00%
SII Payment (Amount)	89.50%

6. References

[1] Zaki M.J., Parthasarathy S., Li W. & Ogihara M., Evaluation of Sampling for Data Mining of Association Rules, *Proc. of the 7th Workshop Research Iss. Data Eng.*, 1997

[2] Agrawal R., Mannila H., Srikant R., Toivonen H. & Verkamo A.I., Fast discovery of associations rules, *Advances in Knowledge Discovery and Data Mining,* 1996

[3] Hand D., Mannila H. & Smyth P., *Principles of Data Mining*, The MIT Press, 2001.
[4] *Clementine 7.0 Users's Guide,* Integral solutions Limited, 2002.
[5] Freitas A, A Genetic Programming Framework for Two Data Mining Tasks: Classification and Generalized Rule Induction, *Genetic Programming: Proc. of the 2nd Annual Conference*, 1997
[6] Han E., Karypis G. & Kumar V., Scalable Parallel Data Mining for Association Rules, *Proc. of ACM SIGMOD Int. Conf. on Management of Data*, 1997
[7] Brin S., Motwani R., & Silverstein C., Beyond Market Baskets: Generalizing Association Rules to Correlations, *Proc. of ACM SIGMOD Int. Conf. on Management of Data*, 1997
[8] Chen M., Han J. & Yu P., Data Mining: An Overview from Database Perspective, *IEEE Trans. On Knowledge And Data Engineering*, 1997
[9] Ikizler N. & Guvenir H.A., Mining Interesting Rules in Bank Loans Data, *Proc. of the 10th Turkish Symposium on Artificial Intelligence and Neural Networks*, 2001.
[10] Ng R., Lakshmanan L. & Han J., Exploratory Mining and Pruning Optimizations of Constrained Association Rules, *Proc. of ACM SIGMOD Int. Conf. on Management of Data*, 1998
[11] Freitas A., A Survey of Evolutionary Algorithms for Data Mining and Knowledge Discovery, *Advances in Evolutionary Computation*, ed. Ghosh A. & Tsutsui S., Springer-Verlag, 2002
[12] Freitas A., On Rule Interestingness Measures, *Knowledge-Based Systems journal*, 1999
[13] Hilderman R. & Hamilton H., *Knowledge Discovery and Interestingness Measures: A Survey*, Technical Report CS 99-04, Department of Computer Science, University of Regina ,1999
[14] Michail A., Data Mining Library Reuse Patterns using Generalized Association Rules, *Proc. of the 22nd Int. Conf. on Software Engineering*, 2000.
[15] Liu J. & Kwok J., An Extended Genetic Rule Induction Algorithm, *Proc. of the Congress on Evolutionary Computation (CEC)*, 2000.
[16] Smyth P. & Goodman R., Rule Induction using Information Theory, *Knowledge Discovery in Databases*, 1991
[17] Fukuda T., Morimoto Y., Morishita S. & Tokuyama T., Data Mining Using Two-Dimensional Optimized Association Rules: Scheme, Algorithms, and Visualization, *Proc. of the ACM SIGMOD Int. Conf. on Management of Data*, 1996.
[18] Srikant R. & Agrawal R., Mining Generalized Association Rules, *Proc. of the 21st Int. Conf. on Very Large Databases*, 1995
[19] Yilmaz T. & Guvenir A., Analysis and Presentation of Interesting Rules, *Proc. of the 10th Turkish Symposium on Artificial Intelligence and Neural Networks*, 2001.

[20] Toivonen H., *Discovery of Frequent Patterns in Large Data Collections*, Technical Report A-1996-5, Department of Computer Science, University of Helsinki, 1996