Application and Diachronic Strength of e-Trans Association Rules

Dr. Vasilis Aggelis
WINBANK (Electronic Banking Division),
Piraeus Bank SA,
Athens, Greece
aggelisv@winbank.gr

Abstract
Investigation of association rules is a very popular data mining method. This method has great importance, particularly in banking sector. Relations between bank products can be identified and analyzed for whole organization activities, or per special banking activity, such as internet banking. In a previous work we introduced e-Trans association rules for Greek internet banking, concerning patterns in e-banking transactions through internet. Those rules were established, taking in account the Greek e-banking market two years ago. Since then, many new electronic services and transactions have been introduced and there is a need to apply e-Trans rules in new datasets. Scope of this paper is the application of those rules, which has two purposes. Firstly is the confirmation of their strength, especially when they have been applied in another bank's datasets than those two years ago. Secondly is the adoption of fresh ones.

Keywords: *Data Mining, e-Banking, Association Rule, A priori, Generalized Rule Induction.*

1. Introduction

Determination of association rules concerning internet banking data is a demanding task since:

- The volume of data is large.
- The goal must be clearly set from the beginning.
- Good knowledge and great experience of the data is a prerequisite not only for the data miner but also for the final analyst.
- Not all rules are of interest.

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

The increase of electronic transactions during the last two years is quite rapid in Greek banking sector. E-banking nowadays offers a complete set of products and services facilitating not only the individual customers but mainly corporate customers to conduct their transactions easily and securely. Discovery of rules between different electronic services is therefore of great significance to a bank.

Identification of such rules offers advantages as the following:

- Good knowledge of the relationships between different types of electronic transactions.
- Description and establishment of most popular internet transactions
- The electronic services become more easily familiar to the public since specific groups of customers are approached, that uses 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.
- Redesign internet transaction structures for those which used rarely
- Reconsidering of the usefulness of products exhibiting little or no contribution to the rules.

This study is concerned with the identification of rules between several different ways of payment. The software used is SPSS Clementine 9.0 Desktop. Section 2 contains association rules' basic features and algorithms, while in section 3 the process of application for e-Trans 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, 15, 18, 19, 21, 27, 28] is given below.
Definition. An association rule is a rule in the form of
$X \rightarrow Y$
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, 15, 19, 21, 27, 28] 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,

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

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 [3].

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, 14] 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, 5, 11, 15, 16, 17], 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.

Although in Clementine [4], the Generalized Rule Induction (GRI) node discovers association rules in the data. GRI extracts a set of rules from the data, pulling out the rules with the highest information content. Information content is measured using an index that takes both the generality (support) and accuracy (confidence) of rules into account.

3. Application of e-Trans association rules

Scope of this study is the application of e-Trans association rules between different payment types internet banking users conduct. Our research deals with payments orders (not standing orders) conducted in Greek e-banking area.

Namely the following types of payments were used:

- Funds Transfer in Other Person's Accounts In house
- Funds Transfer in Other Local or International Bank's Account
- Credit Card Payment Payment of Other Bank's Credit Card
- Funds Transfers in My Own Accounts In house
- Payment of Other Bank's Credit Card
- VAT Payment to Ministry of Economics
- Social Insurance Institute (SII) Payment - Payment Orders for paying employers' contributions
- Power Public Corporation (PPC) bill payment
- Greek Telecommunication Orga-nization (GTO) bill payment
- Credit Card payment
- Allianz General Insurance payment
- Allianz Life Insurance payment
- Vodafone mobile phone bill payment
- Vodafone mobile phone refill online payment
- Cosmote mobile phone bill payment
- Benefaction
- Tellas phone bill payment
- Q Telecom mobile phone bill payment

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.

A sample of the above data set is shown in Table 1.

Table 1: Sample of data set

User	VAT	SII	PPC	GTO	TELLAS	VODAFONE	Q TELECOM	…
…	…	…	…	…	…	…	…	…
User103	F	F	T	T	T	F	F	…
User104	T	T	T	T	F	F	F	…
User105	T	T	T	T	F	T	F	…
User106	F	F	T	T	F	T	F	…
User107	F	F	T	T	F	T	F	…
User108	F	F	T	F	T	F	F	…
…	…	…	…	…	…	…	…	…

In order to confirm e-Trans association rules and discover new ones the A priori [4] and Generalized Rule Induction (GRI) [4] algorithms were used. These rules are statements in the form *if antecedent then consequent*.

4. Experimental Results

Two years ago we obtained the following two most powerful e-Trans Rules [22, 23, 24, 25, and 26]:

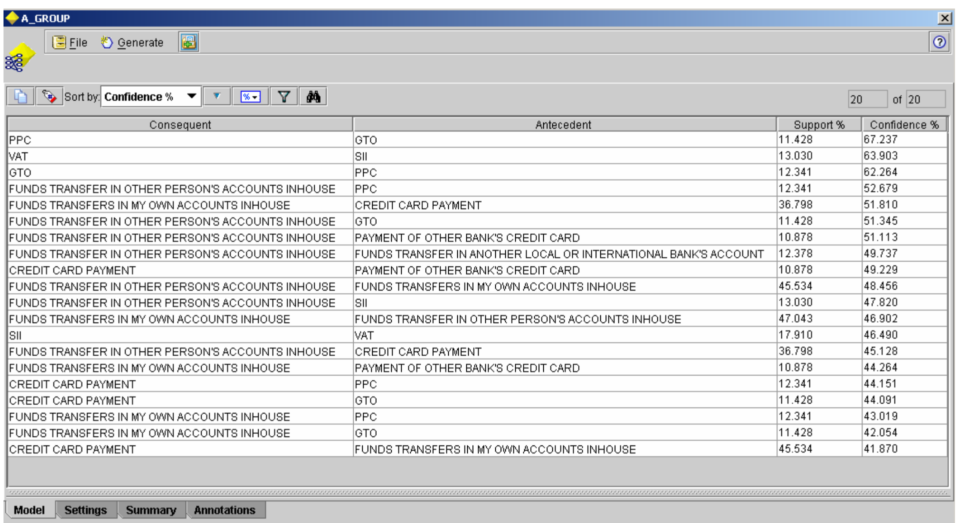
- *If* SII Payment Order *then* VAT Payment Order (confidence=81%). (e-Trans Rule 1)

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

- *If* VAT Payment Order *then* SII Payment Order (confidence= 47%). (e-Trans Rule 2)

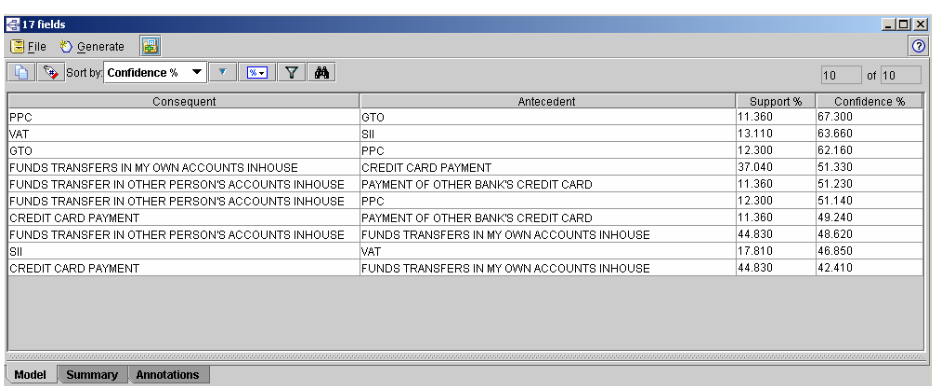
Using A priori method with Minimum Rule Support = 10% and Minimum Rule Confidence = 40%, twenty rules were determined and shown in Figure 1.

Using GRI method with Minimum Rule Support = 10% and Minimum Rule Confidence = 40%, ten rules were determined and shown in Figure 2.



Consequent	Antecedent	Support %	Confidence %
PPC	GTO	11.428	67.237
VAT	SII	13.030	63.903
GTO	PPC	12.341	62.264
FUNDS TRANSFER IN OTHER PERSON'S ACCOUNTS INHOUSE	PPC	12.341	52.679
FUNDS TRANSFERS IN MY OWN ACCOUNTS INHOUSE	CREDIT CARD PAYMENT	36.798	51.810
FUNDS TRANSFER IN OTHER PERSON'S ACCOUNTS INHOUSE	GTO	11.428	51.345
FUNDS TRANSFER IN OTHER PERSON'S ACCOUNTS INHOUSE	PAYMENT OF OTHER BANK'S CREDIT CARD	10.878	51.113
FUNDS TRANSFER IN OTHER PERSON'S ACCOUNTS INHOUSE	FUNDS TRANSFER IN ANOTHER LOCAL OR INTERNATIONAL BANK'S ACCOUNT	12.378	49.737
CREDIT CARD PAYMENT	PAYMENT OF OTHER BANK'S CREDIT CARD	10.878	49.229
FUNDS TRANSFER IN OTHER PERSON'S ACCOUNTS INHOUSE	FUNDS TRANSFERS IN MY OWN ACCOUNTS INHOUSE	45.534	48.456
FUNDS TRANSFER IN OTHER PERSON'S ACCOUNTS INHOUSE	SII	13.030	47.820
FUNDS TRANSFERS IN MY OWN ACCOUNTS INHOUSE	FUNDS TRANSFER IN OTHER PERSON'S ACCOUNTS INHOUSE	47.043	46.902
SII	VAT	17.910	46.490
FUNDS TRANSFER IN OTHER PERSON'S ACCOUNTS INHOUSE	CREDIT CARD PAYMENT	36.798	45.128
FUNDS TRANSFERS IN MY OWN ACCOUNTS INHOUSE	PAYMENT OF OTHER BANK'S CREDIT CARD	10.878	44.264
CREDIT CARD PAYMENT	PPC	12.341	44.151
CREDIT CARD PAYMENT	GTO	11.428	44.091
FUNDS TRANSFERS IN MY OWN ACCOUNTS INHOUSE	PPC	12.341	43.019
FUNDS TRANSFERS IN MY OWN ACCOUNTS INHOUSE	GTO	11.428	42.054
CREDIT CARD PAYMENT	FUNDS TRANSFERS IN MY OWN ACCOUNTS INHOUSE	45.534	41.870

Fig 1. Association Rules determined by A priori



Consequent	Antecedent	Support %	Confidence %
PPC	GTO	11.360	67.300
VAT	SII	13.110	63.660
GTO	PPC	12.300	62.160
FUNDS TRANSFERS IN MY OWN ACCOUNTS INHOUSE	CREDIT CARD PAYMENT	37.040	51.330
FUNDS TRANSFER IN OTHER PERSON'S ACCOUNTS INHOUSE	PAYMENT OF OTHER BANK'S CREDIT CARD	11.360	51.230
FUNDS TRANSFER IN OTHER PERSON'S ACCOUNTS INHOUSE	PPC	12.300	51.140
CREDIT CARD PAYMENT	PAYMENT OF OTHER BANK'S CREDIT CARD	11.360	49.240
FUNDS TRANSFER IN OTHER PERSON'S ACCOUNTS INHOUSE	FUNDS TRANSFERS IN MY OWN ACCOUNTS INHOUSE	44.830	48.620
SII	VAT	17.810	46.850
CREDIT CARD PAYMENT	FUNDS TRANSFERS IN MY OWN ACCOUNTS INHOUSE	44.830	42.410

Fig 2 Association Rules determined by GRI

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

- *If* GTO Payment Order *then* PPC Payment Order (confidence=67%). (Rule 1)

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

- *If* PPC Payment Order *then* GTO Payment Order (confidence= 62%). (Rule 2)

Two other rules obtained exhibiting confidence > 45% are the following:

- *If* SII Payment *then* VAT Payment (confidence=64%), (Rule 3)

valid also reversely.

- *If* VAT Payment *then* SII Payment (confidence=46%), (Rule 4)

Last two rules are same with e-Trans Rules established in 2003 [22, 23, 24, 25, and 26]. So it is confirmed the diachronic strength of them.

As seen, there exists strong relationship between PPC Payments and GTO Payments, as is the case between Funds Transfers and Credit Card Payments. It is important that PPC and GTO Payments were not taken into consideration in the previous work [22, 23, 24, 25, and 26] because they were introduced later in electronic Greek market.

5. Conclusions and Future work

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

The basic outcome is the confirmation of e-Trans rules. VAT and SII Payments are very popular internet transactions and strongly related. Another important outcome is the discovery of fresh e-trans rules. PPC Payments and GTO Payments 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 (households, companies, freelancers) who conduct these payments in order to increase its customer number and profit. Our models 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.

Detection of association, prediction or classification rules can employ several other data mining methods such as decision trees in order to confirm the validity of the results. Clearly, the rules extracted from a decision tree are classification rules, rather than association rules [29].

6. References

[1].M.J.Zaki, S. Parthasarathy, W. Li, and M. Ogihara. “Evaluation of Sampling for Data Mining of Association Rules”, 7th Workshop Research Iss. Data Engg., 1997

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

[3].D. Hand, H. Mannila, and P. Smyth *Principles of Data Mining*. The MIT Press, 2001.

[4].*Clementine 9.0 Users’s Guide*. Integral solutions Limited, 2002.

[5].A. Freitas. “A Genetic Programming Framework for Two Data Mining Tasks: Classification and Generalized Rule Induction”, Genetic Programming 1997: Proceedings of the Second Annual Conference, 1997

[6].E. Han, G. Karypis, and V. Kumar. “Scalable Parallel Data Mining for Association Rules”, Proceedings of ACM SIGMOD Intl. Conf. on Management of Data, 1997

[7].S. Brin, R. Motwani, and C. Silverstein. “Beyond Market Baskets: Generalizing Association Rules to Correlations”, Proceedings ACM SIGMOD Conf. on Management of Data, 1997

[8].M. Chen, J. Han, and P. Yu. “Data Mining: An Overview from Database Perspective”, Ieee Trans. On Knowledge And Data Engineering, 1997

[9].N. Ikizler, and H.A. Guvenir. “Mining Interesting Rules in Bank Loans Data”,: Proceedings of the Tenth Turkish Symposium on Artificial Intelligence and Neural Networks, 2001.

[10].R. Ng, L. Lakshmanan, and J. Han. “Exploratory Mining and Pruning Optimizations of Constrained Association Rules”, Proceedings ACM SIGMOD International Conference on Management of Data, 1998

[11].A. Freitas. “A Survey of Evolutionary Algorithms for Data Mining and Knowledge Discovery”, A. Ghosh and S. Tsutsui. (Eds.) Advances in Evolutionary Computation Springer-Verlag, 2002

[12].A. Freitas. “On Rule Interestingness Measures”, Knowledge-Based Systems journal, 1999

[13].R. Hilderman, and H. Hmailton. “Knowledge Discovery and Interestingness Measures: A Survey”, Technical Report CS 99-04, Department of Computer Science, University of Regina ,1999

[14].J. Hipp, U. Guntzer, and G. Nakhaeizadeh. “Data Mining of Associations Rules and the Process of Knowledge Discovery in Databases”, 2002

[15].A. Michail. “Data Mining Library Reuse Patterns using Generalized Association Rules”, Proceedings of the 22nd International Conference on on Software Engineering , 2000.

[16].J. Liu, and J. Kwok. “An Extended Genetic Rule Induction Algorithm”, Proceedings of the Congress on Evolutionary Computation (CEC), 2000.

[17].P. Smyth, and R. Goodman. “Rule Induction using Information Theory”, Knowledge Discovery in Databases 1991, 1991

[18].T. Fukuda, Y. Morimoto, S. Morishita, and T. Tokuyama. “Data Mining Using Two-Dimensional Optimized Association Rules: Scheme, Algorithms, and Visualization”, Proceedings of the ACM SIGMOD International Conference on Management of Data, 1996.

[19].R. Srikant, and R. Agrawal. “Mining Generalized Association Rules”, Proceedings of the 21st Int'l Conference on Very Large Databases, 1995

[20].T. Yilmaz, and A. Guvenir. “Analysis and Presentation of Interesting Rules”, Proceedings of the Tenth Turkish Symposium on Artificial Intelligence and Neural Networks, 2001.

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

[22].V. Aggelis, “Association Rules Model of e-Banking services” – Transaction on Information and Communications Journal, ISSN: 1743-3517, 2004

[23].V. Aggelis, and D Christodoulakis, “Association Rules and Predictive Models for e-Banking Services” – 1st Balkan Conference on Informatics, 2003.

[24].V. Aggelis, “Data Mining for Decision Support in e-banking area (II)” – 1st International Conference on Knowledge Engineering and Decision Support, 2004.

[25].V. Aggelis, “Association Rules Model of e-Banking services” – 5th International Conference on Data Mining, Text Mining and their Business Applications, 2004.

[26].V. Aggelis, and D. Christodoulakis, “e-Trans Association Rules for e-Banking Transactions” – IV International Conference on Decision Support for Telecommunications and Information Society, 2004.

[27]. Y. Saygin, V. Verykios, and C. Clifton, “Using Unknows to Prevent Discovery of Association Rules, Conference on Research Issues in Data Engineering, 2002

[28]. M. Fernandez, E Menasalvas, O. Marban, J. Pena, and S. Millan, “Minimal Decision Rules based on the Apriori Algorithm”, Int. J. Appl. Math. Comput. Sci., Vol.11, No.3, 2001

[29] A. Freitas, “Understanding the Crucial Differences Between Classification and Discovery of Association Rules – A Positin Paper”, SIGKDD Explorations, Volume 2, Issue 1, 2000.