Making Items Suggestions in Non Online Environments

IOANNIS N. KOURIS*, CHRISTOS H. MAKRIS, ATHANASIOS K. TSAKALIDIS
Department of Computer Engineering and Informatics, School of Engineering
University of Patras
26500 Patras,
HELLAS (GREECE)
&
Computer Technology Institute,
P.O. BOX 1192, 26110 Patras,
HELLAS (GREECE)

Abstract: - Practically all data mining techniques and especially association rules assume an on-line environment in order to operate; such as a web site or an electronic machine. Most businesses though as well as the largest part of current trade are still being made in the traditional off-line way, requiring the physical presence of both the customers and the products, leaving all similar applications unassisted from data mining techniques. In this work we describe a system for delivering items suggestions to customers, tailored for the application that for many initiated data mining research; namely supermarket data.

Key-Words: - Data Mining; Offline Sales; Supermarket Data; Indexing Techniques.

1. Introduction

Since the widespread of computers, database systems and informatics in general, almost every aspect of our every day life has been computerized and digitally recorded. We have in our hands information about the most insignificant (e.g. such as how many records of a very popular artist are sold every hour) to the most important (e.g. the exact rise of temperature in the planet every day). Established companies, scientific organizations, public services agencies and many others have had over the last decades the opportunity to accumulate vast amounts of data. So through the accretion of every day data with past data we find ourselves in a place where we have much more data than we can handle leaving most of it today still unanalyzed. As an example companies such American supermarket "Wal-Mart" reportedly stored about 20 million sales transactions per day [4].

Data mining, formally defined as the extraction of implicit, previously unknown non trivial and potentially useful information from large sets of data, gives us the tools to sift through all these vast data stores in order to find interesting trends, patterns and correlations. Data mining functions include among others: association rule discovery, classification, clustering, summarization, sequential pattern discovery, regression, prediction etc. [1]; [2]. Today, among the most popular techniques in knowledge discovery in databases - KDD, is the extraction of association rules from large database, with applications such as market basket analysis, customer relationship management - CRM, manufacturing diagnosis etc. [3]. Since its first introduction in [11] the task of association rule mining has been one of the most popular and well-studied applications in data mining. The goal of this task is to find all frequent itemsets above a user specified threshold (called support) and to generate all association rules above another threshold (called confidence) using these frequent itemsets as input.

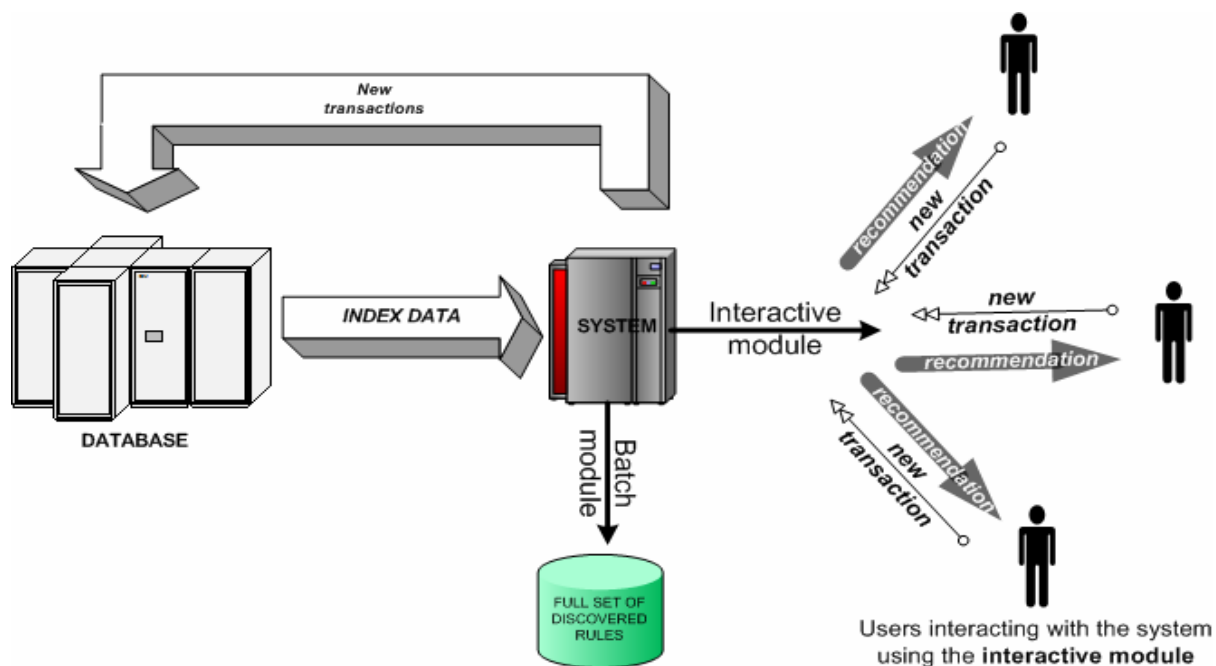* Corresponding author. Tel.: (+30) 6945165529. fax: (+30) 2610429297. E-mail address: jkouris@ceid.upatras.gr

Figure 1: The layout of the indexing mainframe

Clearly the most demanding step is to find all frequent itemsets, since after this step has been completed the generation of all association rules is fairly straightforward. The prototypical application of this task has been the market basket analysis, but the specific model is not limited to it since it can be applied to many other domains (e.g. text documents [12];[13], census data [14], telecommunication data, medical images **Error! Reference source not found.** etc.). In fact any data set consisting of "baskets" containing multiple "items" can fit this model. The use of data mining is not limited to science but has practical interest and applicability in business and industry. Nevertheless commercial organizations and transactional data still gather the largest research effort and interest (not to mention the financial interest). The question is whether we have managed to control and address all the parts of current trade or at least the most important ones? The answer is no since the greatest part of trade and transactions is still being made in an offline manner. The fact is that today the penetration of electronic commerce is not that big and that the use computers and technology by both customers and companies remains comparable fairly low. In fact

The rest of this paper is organized as follows: In section 2 we present the proposed system with all the parts that comprise it. In section 3 we present how the specific system works by giving two typical purchase scenarios. We conclude with section 4 where we give our conclusions.

2. The proposed system

Current techniques and approaches in the field of data mining have been made assuming some form of an online environment. In other words they would operate only through some electronic store of some kind, a web site etc. This is especially true for techniques such as association rules. This of course means that almost all traditional ways of commerce and interaction between physical stores and customers is left unassisted since most data mining techniques cannot be directly applied on them. This suggested a major omission, since despite its radical spread electronic commerce still corresponds only to a small percentage of the sales made in the physical stores. In this work we are proposing the mix of an on-line data mining technique with the wireless technology for addressing a traditionally off line application with immense interest and profits, such as sales in a supermarket. Our system comprises of three main parts which are explained in the subsequent sections, namely:

1. The indexing mainframe.
2. The wireless infrastructure.
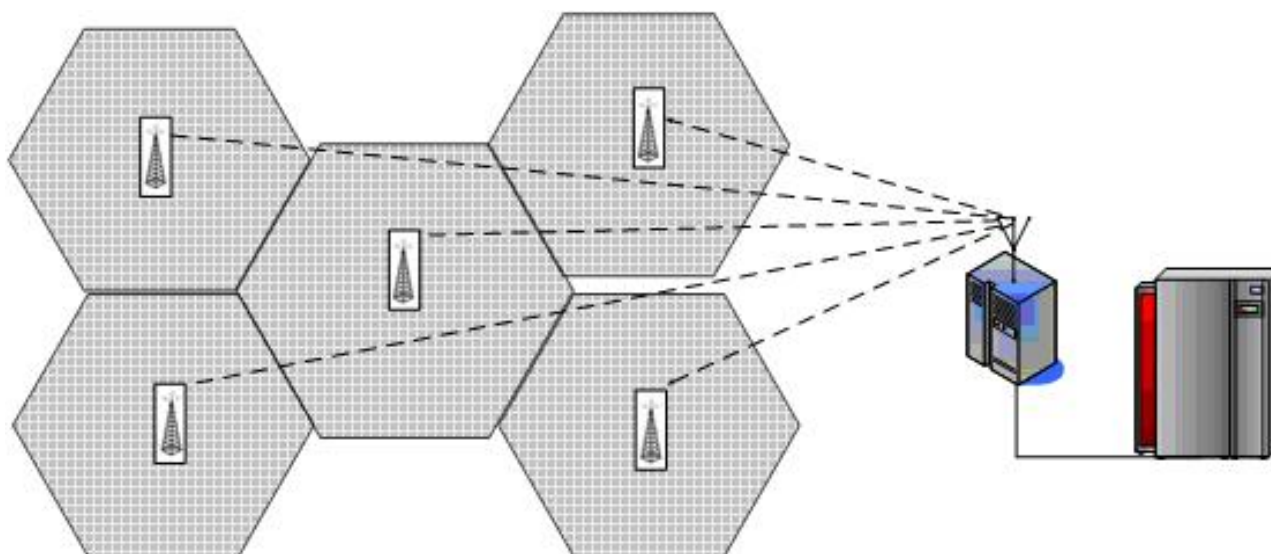3. The shopping carts.

Let's see each part separately.

Figure 2: The wireless infrastructure

2.1 The indexing mainframe

The heart of our system is a structure similar to that proposed in [5], utilizing both modules depending on the type of customer served.

In summary the specific structure takes as input all past transactions stored in our database, generates an index and subsequently is ready to be used in two main ways. Either in its batch module, which

offers services that deviate from the classic scenario used by all other approaches. More specifically instead of going through all the items and their possible combinations and evaluating them using one measure or another, it makes recommendations based on an approach similar to that used when issuing queries to search engines. In practice it tries to answer the following question: Suppose we have a database containing past transactions, and a customer who makes a new transaction. Which transaction(s) best approach the transaction of the specific customer, based on the itemsets already in his basket?

2.2 The wireless infrastructure

The indexing mainframe has a central wireless facility (shown with light blue in Figure 1: The layout of the indexing mainframe

) that handles the connections with the peripheral antennas scattered throughout the store. The antennas are placed strategically inside the store so that they cover all possible areas where a shopping cart might appear. Their placement is dictated by

works exactly like the classic level-wise algorithms (e.g. Apriori) by finding all frequent itemsets and producing all possible rules. The second one, which we will refer to as the interactive module, has a more interactive and on-line behavior. Both modules have as their base an index and work using queries (boolean or ranked), techniques widely used in Information Retrieval. Especially the second module

more or less the same rules that govern the placement of antennas in a cellular phone network[†]. In we can see the whole infrastructure, where the indexing mainframe connects to the central wireless facility that handles all communication to and from the mainframe to the peripheral antennas.

When a shopping cart is in the coverage area of an antenna (shown here as a hexagon) all communication is handled by that antenna. When it moves to another hexagon then all communication is transferred and controlled by the new antenna the same way it happens when a cellular phone moves from one cell to another one. The actual transfer is imperceptible to the user and much easier to be handled by the system, since this is not a critical application such as voice that demands seamless and constant contact.

[†] It is out of the scope of this paper to explain in detail the exact principles of setting up a wireless or cellular network but interested readers may refer to [6][7][8][9][10].

Figure 3: The i-cart

2.3 The shopping carts

The shopping carts are maybe the most difficult, yet the most innovative part of the whole system. In the store there exist a fleet of shopping carts that have an interactive communication with the indexing mainframe through the peripheral antennas. These interactive carts, called i-carts, have three parts attached on them.

1. A bar code reader with a small touchscreen.
2. A memory card reader.
3. A power utility (such as a battery, shown in red), and a small antenna (shown in green bellow).

The bar code reader serves for scanning every item entering our basket. The touch screen provides customers various information such as details about the products, the layout of the store, the route they should follow, the exact location of certain products etc. Also it serves for entering various data and for querying the indexing mainframe (e.g. when was the last time I bought wine, which brand of coffee do I usually buy, which products that might interest me are on special sale today etc.).

The memory card has multiple functions, such as an identification of every customer, a shopping list, an award card etc. It could be any usb memory stick, or any card used by equipment such as digital cameras or mobile phones (such as xd, sd or cf cards). The memory card could be provided by the store, the same way stores provide today customers with award cards. After all the cost of such memory cards is nowadays minimum. Last but not least is the power utility that provides power to the bar code and the memory card reader on the cart and the antenna that handles the communication between the i-cart and the peripheral antennas.

3. A typical purchase

We could identify two possible categories of prospective buyers. Those that are technology savvy, having a computer at home or at work and would like to use the memory card, and those that are more traditional, probably elderly or less educated people that might also cannot afford a computer and would not use the memory card. Let's see what happens with both categories of customers in a typical purchase.

3.1 The technology savvy customers

After walking into the store the customer moves to the i-carts, which are locked. Then he enters his memory card in one of them (the same way we would put a coin in the traditional ones) and that way he unlocks it. The memory card serves first of all as an identification of the customer, whom the system recognizes and greets.

Traditionally customers would usually write down on a piece of paper what they would like to buy and proceed with that information with their sales. Now this task is implemented by the memory card. So the customer would have uploaded on the memory card at home or at his office what he would like to buy. The i-cart reads all products and starts first by suggesting the optimal route for making all the buys as fast as possible. Also it uses the batch module of the central indexing mainframe for suggesting more products to the customer apart from those already in his list. For example if people that buy milk and bread also buy cheese and the specific customer falls into this category but has not cheese among his desired products why not propose him that.

Apart from the products that the customer has written on the memory card to buy, another service provided by the system would be the suggestions made based on historical data on the specific customer. In other words if the system knows for example that the specific customer buys beers approximately every 15 days then it might suggest him to buy beers if the time has arrived and beers are not in his shopping list. The use of the bar code reader in this case is of less use since the system already knows what the customer is going to buy.

3.2 The traditional customers

Even if a customer does not have or does not want to use a memory card he could very well be served by the bar code reader along with the interactive module of the indexing mainframe.

The specific category of customer unlocks and takes also an i-cart but this time with some other way other than the memory card (e.g. by using a coin or a personal identification number typed upon the touch screen). The system does not know who the customer is or what he is going to buy. After having taken the i-cart the customer scans the first product and puts it in his basket. This information is sent through the wireless infrastructure to the indexing mainframe as a query, and the system proposes some other products to the customer. The same procedure is repeated every time a new product is scanned and inserted in the basket, but with ever increasing accuracy in the suggestions made.

4. Conclusions

The specific system is not at all a science fiction scenario and could be implemented almost immediately at any supermarket. Of course there is some significant cost for all the required equipment and software that could only be justified if the investment is made on a large supermarket. Nevertheless this cost would be amortized fast enough by the extra sales made and the new customers that would be attracted.

The scenario using memory cards is far more certain to be used since the method utilizing the use of the bar code readers is not at all sure. Nevertheless the customers could be "forced" (e.g. by using some form of cover on the cart that would only open after having red the bar code of some product) or "encouraged" (e.g. by providing added value information of the product to the customer such as ingredients, land of origin, nutritional value etc.) to use it.

5. Aknowledgements

6. References

[1] P. Adrians and D. Zantige. Data Mining. Addison-Wesley, 1996.

[2] M.S. Chen, J. Han and P.S. Yu. Data Mining: an overview from a database perspective. *IEEE Transactions on Knowledge and Data Engineering*, vol. 8, no. 6 (1996) pp. 866--833.

[3] U. Fayad, G. Piatetsky-Shapiro and P. Smyth. From Data Mining to knowledge discovery: an overview Advances in Knowledge discovery and data mining. Cambridge, MA, USA: AAAI Press/The MIT Press, 1996.

[4] S. Hedberg. The data gold rush. In *BYTE*, pp. 83--88, October 1995.

[5] I.N. Kouris, C.H. Makris and A.K. Tsakalidis. "Using Information Retrieval techniques for supporting data mining". In Data & Knowledge Engineering, Vol. 52(3), 353—383, March 2005.

[6] MacDonald V. H.:"The cellular concept", Bell System Technical Journal, Vol. 58, No 1, Jan 1979, pp.15--41.

[7] Jakes W.C.: "Microwave Mobile Communications", New York, John Wiley Publications, 1974.

[8] Henry P.S. and Glance B.S.:"A new approach to high capacity digital mobile radio", B.S.T.J., 60, No. 8, 1981, pp. 1891--1904.

[9] Wisniewski S.: "Wireless and Cellular Networks", Prentice Hall, 2004.

[10] Mishra A. R.: "Fundamentals of Cellular Network Planning and Optimisation : 2G/2.5G/3G... Evolution to 4G (Hardcover)", John Wiley & Sons, 2004.

[11] Agrawal, R., Imielinski, T. and Swani, A., 1993, 'Mining association rules between sets of items in very large databases', Proc. ACM SIGMOD Conf. Management of Data, pp.207-216.

[12] Brin, S., Motwani, R, Ullman, J.D. and Tsur, S., 1997, 'Dynamic Itemset Counting and Implication Rules for Market Basket Data'. Proceedings of the ACM SIGMOD, pp. 265-276.

[13] O. R. Zaine, J. Han, and H Zhu. Mining recurrent items in multimedia with progressive resolution refinement. In Int. Conf. on Data Engineering (ICDE 2000), pp. 461-470, San Diego, CA, February 2000.

[14] J. D. Holt and S. M. Chung, ``Mining Association Rules Using Inverted Hashing and Pruning,'' Information Processing Letters, Vol. 83, No. 4, Elsevier Science, 2002, pp. 211--220.

[15] J. D. Holt and S. M. Chung, ``Multipass Algorithms for Mining Association Rules in Text Databases,'' Knowledge and Information Systems, Vol. 3, No. 2, Springer-Verlag, 2001, pp. 168-183.