Symmetric Adaptive Customer Modeling in an Electronic Store

Maria Barra Giuseppe Cattaneo Alberto Negro Vittorio Scarano

Dipartimento di Informatica ed Applicazioni "R.M. Capocelli"
Università di Salerno, 84081 Baronissi (SA) – Italy

1. Introduction

Electronic Commerce (EC) is currently one of the fastest growing and most practically relevant application areas of distributed systems technologies. It is based on the economic aspects of commercial trading patterns combined with distributed computing systems technology. It is a market environment that is characterized by low transaction costs, a large number of market participants, and easy online access to services and goods offered. It also implies a set of rules and policies for the successful organization of business transactions.

EC involves more than simple online transactions, it encompasses diverse activities as conducting market research, identifying opportunities and partners, cultivating relationships with customers and suppliers, document exchange and customer modeling.

Our paper deals with the latter aspect of EC. We introduce here a model for developing a symmetric adaptive system for EC on the World Wide Web. Our main contribution is that the model is, by all means, symmetric: we model both customers and goods and make both their profiles change as a consequence of a customer buying a certain product. The symmetry in our model greatly simplifies the approach and the queries, giving some insights on the formalization of the allowed queries that were, in way, unexpected. Furthermore, the model itself can provide an easy-to-evaluate measure for the confidence in adapting its response to any given customer and is able to provide useful feedback to the manager, then allowing, so to speak, "manual adjustment" that can help the behaviour of the system in the future.

Our model is simple, easy to implement, not computationally expensive, and can be used as a base to build a more refined system upon. We will show, just as a proof of concept, how several kind of queries (with relative confidence measure) can be obtained in the model. We emphasize that we see as a strength of the model the fact that several other possible kind of queries (more refined and complex) can be easily obtained in the model. One of the main challenges that we faced was to keep the model simple and, at the same time, extensible. A brief description is in the final section of concluding remarks.

We anticipate here that our focus is not on the electronic payment transactions but in what goes on before it, i.e., the usual information exchange between the customer and an *electronic clerk*, attending the desk of our *electronic shop*.

1.1. EC Systems and World Wide Web

In the last few years, many systems have been proposed and implemented that are meant to offer capabilities to support Electronic Commerce. Here, we try a first categorization of these systems by using a real-world analogy of a shop that has a Window, where potential customers can see goods, has a Clerk that is able to intelligently interact with customers, sell them goods and, finally, has a Cash where customers pay (either with cash, checks of cards).

Our interest is on Clerk Systems that are receiving far less attention than Electronic Cash Systems. Our goal is to design a mechanism that allows for interactive, intelligent, guided and adaptive shopping that is certainly more fruitful both from the customer's perspective (finding interesting items in less time) and from the merchant's perspective (being able to respond to customer's unexpressed desires).

Although originally developed to facilitate information sharing (especially in academic environments), the World Wide Web has large potentials, partly already achieved, in the field of Electronic Commerce.

It is our opinion that in order to fully exploit WWW potentiality (user interface, portability, extensibility and widespread access) in the EC field, the merchant server must be a modified Web server, that must be aware of user's (the customer) behaviour so that it can take into account his interests and, as a consequence, provide the user with the right products, in few words it must be adaptive.

The problem, in a certain way, is similar to those experienced when WWW is used as an educational support, especially for distance learning. Several educational systems based on WWW have been developed in the recent past [3, 6] and some of them try to introduce adaptivity as a powerful tool to provide a better response, see [5] for a detailed bibliography.

In general, such adaptivity mechanism can be helpful to avoid the risk that users can "*have trouble in finding the information they need*" [7] because of the large amount of information available. An adaptive response from server can be able to avoid the "*information overloading*" by slowly increase the amount of information given to the user, information that is carefully chosen given the knowledge of the user.

2. The model

In this section we describe the model in a simplified scenario where a single commercial server is available and several customers access the server. This scenario is for ease of description but is not totally unrealistic. The model in a distributed environment is described in the companion paper [2].

We assume that customers are divided in categories, as well as items. When a new customer arrives, or a new item is available for selling, it is possible to assign it a *profile* corresponding to a stereotype that is chosen at the beginning. Stereotypes can be, of course, different and chosen as a consequence to the customer/item belonging to a category.

Informally, the idea of the profile of a customer is to define a category and a vector of values that indicate *his* current predisposition to buy goods from certain products categories. Symmetrically, an item of merchandise has a category and a vector of values that indicate *its* predisposition to be bought from a customer of each category. The symmetry here is an innovative aspect of the method: we can develop queries that go both ways and that, in our opinion, are a surprisingly formalization of common practise in the EC field.

We propose, here, also, several queries that can be performed on such a model and that can be used either to adapt the response toward customers but also to adjust the behaviour of the system by providing a useful feedback to the manager. For example, critical to the convergence of the system toward good responses is choosing right stereotypes and categories. Feedback can indicate that some stereotypes are too crude or that some categories have to be split in half or two categories joined.

We describe the requirements that are to be obtained by our model. This requirements are expressed in terms of easy and efficient management from the merchant side. The model:

- must model customer's behaviour: if a customer interacts with the system, the status of the system after the interaction should depend on the interaction;
- must be symmetric. Queries available for customer modeling should be also usable for merchandise modeling,to help the store manager;
- must be able to measure the *surprise* of a particular customer buying a certain item of merchandise in order to provide feedback to the store manager; i.e., this could trigger an *alarm* for the manager of the electronic store, maybe the prize is too low, or maybe it is time to buy more items of the same kind (since they are particularly appreciated by people of an unexpected category) or maybe the customers' stereotypes are surpassed by the society and it may suggest that some stereotypes are too crude or plainly wrong;
- must be able to provide a measure of the confidence of the mechanism in the forecast of the customer's behaviour;
- must be able to characterize *groups* of homogeneous customers or items. That is, we would like to get a sample of customers whose behaviour is similar enough, maybe in order to propose them a "special offer". Or, on the other side, we would like to group *close* items (i.e. for example, directed to the same audience) in such a way to lower prizes for them. Or, maybe even more important, start from a set of homogeneous customers, then observe them and propose to one of them what the other has recently bought.

At the same time, there are characteristics that we would not like in our model, such as follows:

- We do not want to keep history information: each customer can have some field where special offers are selected for them ad offered as soon as they get connected.
- We do not want computationally costly operations to be performed on the model: our system can have as many customers as millions and (for example) averaging over all the customer is simply unacceptable.

2.1. Queries in our model

In the rest of the paper, we call *Customer* any buyer that is connected to a commercial site (called *Merchant*).

The queries that we would like to be answered by our model are divided in two class: we show now queries in the interaction Customer/Merchant.

- *CUSTOMER: What is interesting to me?*

List of the "top ten" products for a given customer (with a measure of confidence in the forecast).

- *MERCHANT: Who can be interested in this?*

List of the "top ten" customers for a given item of merchandise (with a measure of confidence in the forecast).

• *CUSTOMER: Which merchandise group does this item belong to?*

What are the goods that show a certain degree of affinity with this particular good I just bought (with a degree of confidence in the forecast).

• *MERCHANT: Which customers group does this customer belong to?*

What are the customers that show a certain degree of affinity with the customer that just bought this good (with a degree of confidence in the forecast).

Now we describe the queries that are related to system management and marketing policies. They are meant to provide effective and efficient *feedback* to the management.

• *MERCHANT: Which confidence do I have in profile of this customer/good?*

The merchant should provide a measure of behaviour of the system on customers/goods and of its confidence.

• *MERCHANT: Should I change the stereotypes?*

if all (or a large percentage of) the customers of the same category move together toward a *limit* configuration, maybe this should be the stereotype. The same for items of merchandise.

• *MERCHANT: Should I change the categories?*

If all the customers of a given category split in exactly two subcategories (with relative limit configurations) maybe I should split the categories as well. The same for items of merchandise.

Moreover, it is easy to set an *alarm* that when the *surprise* of a transaction (the term is defined shortly) is above a threshold is able to asynchronously call the manager to inform him that there is something strange.

2.2. The theoretical model

As we said at the beginning of section 2, the profile of a customer consists of his category and a vector indicating the degree of interest (probability values) that the customer has with respect to each kind of merchandise. The same can be symmetrically said for goods' profiles. We now define formally the profiles and the model.

For x, y positive integers, let $A(x,y)$ be a set

$$A(x,y) \subseteq \{1,\ldots,x\} \times [0,1]^y.$$

An item $\mathbf{a} \in A(x,y)$ is a pair $(c, \langle a_1, a_2, \ldots, a_y\rangle)$. The first component of $\mathbf{a}$, denoted by $\mathrm{Cat}(\mathbf{a})$, is called the *category* while the remaining y-item vector is called its *configuration*. A configuration is *legal* if $\sum_{i=1}^{y} a_i = 1$.

The categories implicitly define a partition of $A(x,y)$ in x subsets $A_1(x,y)$, $A_2(x,y)$, …, $A_x(x,y)$ where

$$A_c(x,y) = \{\mathbf{a} \in A(x,y) \mid \mathrm{Cat}(\mathbf{a}) = c\}.$$

The *symmetric adaptive* model is composed by two sets:

• the set of customers $U = A(d,k)$ where customers are partitioned in *divisions* $U_1, U_2, \ldots, U_d$;

• the set of merchandise $M = A(k,d)$ where items are partitioned in *kinds* $M_1, M_2, \ldots, M_k$.

A customer $\mathbf{u} \in U$ is, therefore, characterized by his *division* $\mathrm{Cat}(\mathbf{u})$ and by a configuration vector that has as many components as the kinds of merchandise available. Analogously, a merchandise item $\mathbf{m} \in M$ has its own *kind* $\mathrm{Cat}(\mathbf{m})$ and its configuration vector has as many components as the number of divisions.

Let us define the *surprise* of a customer $\mathbf{U} = (c, \langle u_1, \ldots u_k\rangle)$ related[1] to an item of merchandise $\mathbf{M} = (s, \langle m_1, \ldots, m_d\rangle)$ as the function

$$S(\mathbf{U},\mathbf{M}) = 1 - u_s m_c$$

where u_s and m_c represent the elements' values of the profiles of user and good that are related.

This function measures the information that is given to the system when $\mathbf{u}$ buys $\mathbf{m}$ and goes from 0 (absolutely expected) to 1 (maximum surprise).

The function *distance* is the usual euclidean distance between points in $\{0,1\}^k$ for customers and in $\{0,1\}^d$ for items of merchandise.

2.3. The operations

• *Buying a product and updating profiles.*

What happens to our model when customer $\mathbf{U} = (c, \langle u_1, \ldots, u_k\rangle)$ wants to buy item $\mathbf{M} = (s, \langle m_1, \ldots, m_d\rangle)$? Let

$$P = \max\{S(\mathbf{U},\mathbf{M}), u_s\} \ \ \textit{and} \ \ Q = \max\{S(\mathbf{U},\mathbf{M}), m_c\}.$$

Then, in $\mathbf{U}$ the value P takes the place of u_s and in $\mathbf{M}$ the value Q takes the place of m_c and this means that either configurations do not change if the event is not a surprise, that is $S(\mathbf{U},\mathbf{M}) \le u_s$ or $S(\mathbf{U},\mathbf{M}) \le m_c$. If a profile changes due to a surprise, it has to be made legal (i.e. summing up to 1) and, therefore, assuming that $U' = (c, \langle u'_1, \ldots, u'_k\rangle)$ and $M' = (s, \langle m'_1, \ldots, m'_d\rangle)$ are, respectively, the customer and good profiles after $U = (c, \langle u_1, \ldots, u_k\rangle)$ has bought the item of merchandise $M = (s, \langle m_1, \ldots, , m_d\rangle)$ we have that:

$$u_i = \begin{cases} \frac{1-u_s \cdot m_c}{C} & \text{if } i = s \\ \frac{u_i}{C} & \text{otherwise} \end{cases}$$

$$m'_i = \begin{cases} \frac{1-u_s \cdot m_c}{D} & \text{if } i = c \\ \frac{m_i}{D} & \text{otherwise} \end{cases}$$

where $C = \sum_{i \neq s} u_i + (1 - u_s m_c)$ and $D = \sum_{i \neq c} m_i + (1 - u_s m_c)$

[1] We want to emphasize by the term "related to" the bidirectionality of the system. We do not say that $\mathbf{U}$ *buys* $\mathbf{M}$.

Notice also that, the preference of a certain customer U for items of kind i, given by u_i, can be increased due to the surprise of U buying an item of kind i that is not usually sold to people of the same category of U. The value u_i can be also decreased due to surprises in other fields that make necessary make the configuration legal (i.e. summing up to 1).

- *"Top 10" queries.*

Here we describe the "Top 10" query for a customer, the analog query for items of merchandise is totally symmetric.

Assume we want to present T items of interest for customer $\mathbf{U} = (c, \langle u_1, \ldots, u_k \rangle)$. Such a list is built with the following proportions: there are as many as $u_i \cdot T$ items[2] for each kinds of goods i.

What are the goods in kind i selected for "top T" products for customer $\mathbf{U}$? We select the items that are most wanted by customers in category c, that is we pick the goods that have maximum m_c.

We emphasize here the advantages of the symmetric model: queries are easily obtained by reverting things in the model.

- *Affinity queries.*

Because of space limitation, we sketch here how affinity queries can be performed in our system. Recall that the second component of customer profile is a vector with k elements. The distance function used is the usual euclidean distance in the sphere S with radius 1 and center t the origin in the space $\{0,1\}^k$. Now, an affinity query can be easily seen as a query of the kind: given a point U in sphere S, find all the points within distance ℓ from U. The same is done for items of merchandise in $\{0,1\}^d$.

2.4. Confidence of the system

In this subsection, we show how our model can be used to give a measure of degree of confidence in the operations previously described. This can be used to provide useful feedback to the manager.

We describe here how to evaluate the confidence that can be assumed for customers' profiles. The confidence for goods' profiles can be easily obtained by clerical changes, since the symmetry of our model.

Let us define the entropy function of a customer U and of an item of merchandise, that are random variables respectively as follows[3]:

$$H(U) = -\sum_{i=1}^{k} u_i \log u_i \quad and \quad H(M) = -\sum_{i=1}^{d} m_i \log m_i.$$

[2] At this moment we are disregarding rounding problems. Solutions "*ad hoc*" can be easily adopted.

[3] All the logarithms in the rest of the paper are base 2

In the sequel, we give a very brief definition and interpretation of the entropy as defined by C.Shannon [8]. The interested reader may consult [1]. Entropy function is meant to measure the average uncertainty of a random variable X. $H(X)$ does not depend on the values the random variable assumes, but only on probabilities associates with those values. In our model, U represents the probability distribution of the customer buying products of a certain category. The same interpretation can be given of the item of merchandise M.

The entropy of a random variable with ℓ values ranges from 0 to $\log \ell$. The smaller the values of $H(X)$ are, the more certain is the outcome of the random variable X. In our model, if entropy of a customer U approaches to $\log k$ then it means that customer U is, more or less, buying items from all the kinds.

Now, what we want is to evaluate $H(U')$ as a function of U where U' is the customer profile after U has bought the item of merchandise M and where U M and U' are defined as in 2.2

2.4.1. Operations Confidence

Let us consider the relation between the interactions number of a customer and her entropy. Let i be the number of interactions of customer U, $H(U)$ his entropy and M the maximum number interactions that a customer had with the system. Let

$$N(U) = \left(\frac{i}{M}\right) \log k$$

be the ratio of the number of interactions scaled up to $\log k$. Recall that $\log k$ is the maximum of the entropy $H(U)$ and is achieved for equally distributed profile.

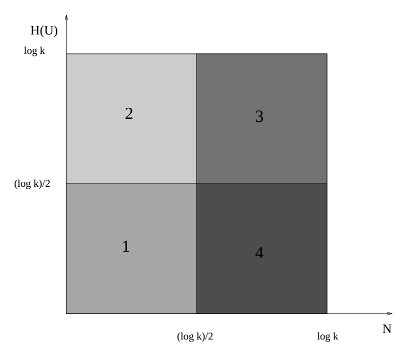

Figure 1: A diagram of the four regions where the point $(N(U), H(U))$ is located. The different shadows in each region show the expected number of customers that a well-behaved system should have.

Each interaction moves the point $(N(U), H(U))$ into a 2-dimensional space. If we analyze this space: we consider it divided in four regions. Each pair $(N(U), H(U))$

must fall into one of them. We know that $0 \leq H(U), N(U) \leq \log k$.

• If $N(U) \leq \log k/2$ and $H(U) \leq \log k/2$ we say that point $(N(U), H(U))$ falls in the area 1 (see Fig. 1). It means that customer U shows some particular tendencies to buy goods in certain categories, but confidency is not high due to the few interactions of U with our system.

• In area 2, ($N(U) \leq \log k/2$ and $H(U) \geq \log k/2$) we have customers that have shown (during few interactions) a behaviour that makes any guess a "wild guess" since his profile could be just the equally distributed stereotype assigned by some policy of the system.

• In area 3, ($N(U) \geq \log k/2$ and $H(U) \geq \log k/2$), we have customers that are consistently showing an equally distributed behaviour. This means that even small deviations from the average value, $u_i = \frac{1}{k}$ can be significant and can be trusted.

• In area 4, ($N(U) \geq \log k/2$ and $H(U) \leq \log k/2$), we have customers that are really well classified, i.e. our system is able to guess with a substantial confidence the next moves of the customer.

How do we use the confidence in the queries previously defined? Every time a customer U is given back as a result of the query, the system sends the confidence (defined as the pair $(N(U), H(U))$) along with the results. The confidence can be used as a numerical measure to guide into the interpretation of the results. For example, low confidence in a customer given back from a "Top 10 customers" query. Recall that this query is needed, for example, when a new item is arriving into the system and the manager wants to find the best potential customers[4]. Now, a low confidence given to a customer U means for the manager, that, for example, customer U had only few interactions with the system and the fact that he is in the "Top 10 customer" for the given item may depend on his stereotype (the one assigned by the system at the entrance) and not by his behaviour that is not well known by the system.

The figure 1 shows the regions just defined. The idea is that in region 4 there is the majority of the customers: those that have interacted a lot and, at the same time, that have a "stable" profile (low entropy). In region 3 are placed the customers that are consistently buying products form all the kinds (high entropy) and therefore difficult to "*have a snapshot*" of. Notice also that feedback to the manager can help in moving "average" value (arbitrarily fixed by us in half the maximum value $\log k$) back or forth on the axis so as to divide and subdivide different regions.

A possible other usage of such a confidence is, for example, changing the effect of some operations if the query is originated by a customer in region 1 or 2. For example, a "Top 10 items" for a customer with few interactions could be just a "default" value.

[4] It can be also used for special offers.

3. Conclusions and Further research

One of the greatest advantages of our model of commercial service is that it is simple enough to be easily described, implemented, tested and evaluated and, it is opening new avenues for more refined kind of queries; it is also extensible to support security protocols and policies needed in an untrustworthy environment [2], where privacy and attackers must be considered.

One possible improvement, for our model of commercial service, is on the feedback to the manager section. For example, an accurate statistical analysis on the field can decide to develop more refined *trigger* mechanisms, such as deciding to discard (as statistical not relevant) exceptionally high values of surprise.

Another improvement to the behaviour of the model involve to change the surprise as having different weight depending on the areas (1,2,3 or 4) where the customer/items belongs to.

References

[1] R.B.Ash. "*Information Theory*". Dover Publications,Inc. New York 1965.

[2] M.Barra, G.Cattaneo, M.Izzo, A.Negro, V.Scarano. "*Symmetric adaptive Customer Modeling for Electronic Commerce in a Distributed Environment*". Proc. of Intern. IFIP Working Conf. on "Trends in Distributed Systems for Electronic Commerce", LCNS Springer-Verlag Eds. Hamburg (Germany), June 3-5, 1998.

[3] D. Dwyer, K. Barbieri, H.M. Doerr. "*Creating a Virtual Classroom for Interactive Education on the Web*". Proc. of WWW 95, Third Int. Conf. on World Wide Web.

[4] D.Cameron. "*Electronic Commerce. The New Business Platform for the Internet* ". Computer Technology Research Corp. 1997.

[5] S. Ferrandino, A. Negro, V. Scarano. "CHEOPS: Adaptive Hypermedia on World Wide Web". Proceedings of the European Workshop on Interactive Distributed Multimedia Systems and Telecommunication Services (IDMS '97), 10-12 Sett. 1997. Ed. Springer-Verlag (LNCS).

[6] B. Ibrahim, S.D. Franklin. "*Advanced Educational Uses of the World Wide Web*". Proc. of WWW95, 3rd International Conference on World Wide Web.

[7] J. Nielsen. "*Hypertext and Hypermedia*". Academic Press Ltd, 1990.

[8] C.E. Shannon. "*A Mathematical Theory of Communication*", *Bell System Tech. J.*, **27**, pp.379-423, 623-656.