User Profiling with Hierarchical Context: An e-Retailer Case Study

Cosimo Palmisano[1], Alexander Tuzhilin[2], and Michele Gorgoglione[1]

[1] Politecnico di Bari, Management Dept.,
c.palmisano@poliba.it m.gorgoglione@poliba.it
[2] New York University, Stern Business School, IOMS Dept.,
atuzhili@stern.nyu.edu

Abstract. In e-commerce applications, no systematic research has been provided to evaluate if the use of a detailed and rich contextual representation improves the user modeling predictive performances. An underestimated issue is also evaluating if context could be inferred by existing customer data off-line, in spite of getting the customer involved on-line in the gathering process. In this paper, we address those problems, defining context as "the intent of" a customer purchase. To this aim, we collected data containing rich contextual information, hierarchically structured, by developing a special-purpose browser. The experimental results show that the finer the granularity of contextual information the better is the modeling of customers' behavior. Representing the context in a hierarchical structure is a necessary condition, for inferring the context off-line, but it's not a sufficient one.

Keywords: E-commerce, context hierarchy, context inference.

1 Introduction

In his interview at AMA (American Management Association) P.K. Prahalad [25] stated that *"the ability to reach out and touch customers anywhere at anytime means that companies must deliver not just competitive products but also unique, real-time customer experiences shaped by customer context"* and that this would be the next main issue for the CRM practitioners. There exists substantial anecdotal evidence in the press and the popular literature that supports Prahalad's observation, including scientific literature in the field of customer profiling and recommendation systems. For example, the director of personalization at one of the major on-line retailing companies once received a nasty email from the CEO telling him that he should either fix his personalization system or lose his job. The CEO's email was prompted by a customer's complaint that the company's personalization system was making offensive assumptions about the lifestyle of this customer and was recommending inappropriate products to that person. Upon a closer examination, it was discovered that the customer once bought an item as a gift for his friend, and the personalization system started recommending related products to that customer making implicit assumptions about his lifestyles, which infuriated that customer. This true story is

B. Kokinov et al. (Eds.): CONTEXT 2007, LNAI 4635, pp. 369–383, 2007.

very symptomatic of problems pertaining to many personalization systems that often predict customer behavior from the registration and the purchasing information of online customers without studying the contexts in which these purchases are made. In the previous example, if the system knew that the purchase was made in the context of a gift, this transaction should have been discarded from inferring that customer's behavior, and the whole problem would have been avoided. Getting such contextual data characterizing the circumstances in which purchasing or other on line transactions took place, such as the "intent of" a purchase, special payment conditions, economic climate and the customer's geographic location, is not easy in many e-commerce applications. For instance, it may not be practical to ask the customer about the purpose of his/her purchase because of privacy and some technological constraints. Therefore, before acquiring such contextual information, it is necessary to provide hard scientific evidence that this contextual information indeed makes a significant difference in building better customer models. In our prior work [12], we addressed this problem and demonstrated that (a) contextual information matters in the sense that it facilitates building better personalized predictive models of customer behavior, and (b) granularity of contextual information also matters, i.e., the more granular and the more specific the contextual information is, the better we can predict customers' behavior. However, the contextual information usually does not come as a set of various alternatives, such as buying a product for yourself or as a gift. Usually, it is organized in the form of a context hierarchy, where coarser types of context are partitioned into progressively finer levels of contextual information.

In this paper, we study the questions of whether (a) it is feasible to infer these whole hierarchies of context from the data and, (b) whether the inferred hierarchical models outperform individual-level models of contextual knowledge. We show that the more we know about the context of a transaction, the better we can predict the customer's behavior. We also show in our experiments that it is possible to infer the contextual information of a transaction with a reasonable degree of accuracy.

Answering these questions has relevant managerial implications. In fact, acquiring contextual knowledge is costly in terms of privacy and technological issues and this cost is even higher if this knowledge is represented by complex structures. Providing a systematic analysis about this research issue is important to evaluate what is the better approach to make this knowledge useful and the gathering process of rich contextual knowledge as worth as possible.

2 Literature Review

In our previous work on context [12] we addressed the problem of investigating if contextual information indeed makes a significant difference in building better customer models in e-commerce applications. To this aim we collected experimental purchasing data of customers and the "intent of purchase" was gathered as contextual information. The overall contextual purchasing option was defined by a tree-shaped hierarchical structure where the root was the coarser representation of the contextual information and the leaves the finer knowledge representation. After collecting all the purchasing data, we built predictive models of purchasing behavior for the contextual and un-contextual cases under different experimental settings. The more relevant settings for our research purposes were: the degree of contextual information and the granularity of customer segments. In the first case the aim was to evaluate how the prediction performances change at different levels of contextual knowledge. In the second one the aim was to evaluate at which unit of analysis we get better predictive performances. In this prior work has been demonstrated that (a) contextual information matters in the sense that it facilitates building better personalized predictive models of customer behavior, and (b) granularity of contextual information also matters, i.e., the more granular and more specific the contextual information is, the better we can predict customers' behavior.

The basic hypothesis of the previous work was that contextual information is available and ready to be used to label each transaction; but it's not always the case. In many situations context could not be easily available and its gathering can be too expensive because of privacy concerns and various other considerations. In those cases, one possibility would be to infer the context from existing data off-line, thus reducing the costs of collecting it on-line, and avoiding user intervention in data collection. This issue was not investigated in our previous work and will be investigated in this paper. In particular the differences in performance between inferring a single level of context and inferring the whole hierarchy of contextual knowledge will be systematically evaluated.

Scholars in marketing have maintained that the purchasing process is contingent upon the context in which the transaction takes place. The same customer can adopt different decision strategies and prefer different products or brands depending on the context [6], [17], [22]. According to [20], "consumers vary in their decision-making rules because of the usage situation, the use of the good or service (for family, for gift, for self) and purchase situation (catalog sale, in-store shelf selection, and sales person aided purchase)." Therefore accurate prediction of consumers' preference undoubtedly depends upon the degree to which we have incorporated the relevant contextual information. Those statements have been also supported by the results of our previous work. The importance of including contextual information in recommender systems has also been demonstrated in [2], where a multidimensional approach to recommender systems is presented. In general, it is possible to assert that the ability of exploiting the knowledge of context is expected to increase the potential of many applications aimed at delivering services to users [1]. Other contributions to the context paradigm have been provided in information retrieval [7] [19], web browsing personalization systems [13], [31], Web services [23]. Most of this work has also tried to determine how to improve the use of context by applying different representations of contextual knowledge using taxonomies, hierarchical structures model or semantic web appliances. In [7] Bothorel and Chavalier propose the click stream data as an identification criterion and create a rich context for providing clues to recommend relevant web pages links to an unknown user. In [13] the use of Semantic Web is suggested for facilitating the capture of knowledge regarding users' context, and supporting the performance of web searching tasks. In [31] Zhu et al. introduced the notion of structured contexts and show its effectiveness using a lightweight ontology to provide a structure for representing contexts in an online price comparison example.

In e-commerce, the concept of a context is associated with Contextual marketing which is the strategy of providing personalized information (advertisements, banners, offers) to customers at the point of need in real time, [16], [21].

If the concept of customer and that of transaction are broadened to embrace any user interacting with a company or an application to get a service, then the importance of knowing the context is recognized in other fields and applications. For instance, context-aware systems are designed to exploit the contextual information available to better serve the user [9], and to adapt to changes in the context. In [24] is introduced the concept of primitive context as the basic context abstraction for formalising and reasoning about context in a consistent and conceptually simple way.

All the web and context-aware examples cited have the common characteristics of reasoning about situations in which contextual knowledge exists, is reliable and has been represented in different structures. But the reliability of contextual data in certain industrial situations is very expensive in terms of technological investments and privacy concerns. E-commerce is a typical example of a sector where privacy concerns and technological constraints are really high. Therefore, it is important to determine how strategic it is to acquire the contextual information, when modeling the behavior of a customer [8], [9]. So far scholars, belonging to different fields, have addressed the problem of inferring the contextual data off line, in spite of getting the user involved in the gathering process. One approach to context recognition and inference is based on supervised learning which requires the intervention of an expert, or the user, at some point of the process to label contexts or define the user needs in a given context. A second compelling opportunity is setting up an unsupervised learning stage to learn associations between contexts and user needs without explicit user intervention. An example of contextual knowledge inference has been studied in [28] by the concept of "granularity". They demonstrate increasing the level of granularity of spatial and temporal context data tends to provide good inferential properties in their natural language processing application. Another example of contexts inference is applied in text documents [11], where the aim of the work is to infer context taxonomy for locating the right documents by using contextual indexing or contextual reasoning. Other researchers have applied a comparative approach in order to experimentally evaluate if the inference problem is feasible in an unsupervised way or not. In [10] an unsupervised learning approach to context recognition has been compared to supervised models. Another example of comparison between contextual inference using supervised and unsupervised technique is provided in [26]. This paper describes how probabilistic graphical models learned with different Acyclic Directed Graphs could exploit context represented as statistical dependences. In [26] a supervised approach developed by the expert is useful to elaborate the more efficient Bayesian Network for improving predictive performances. The main statement is that even if many unsupervised algorithms for drawing more efficient probabilistic graphical models are available, they still require an assistance of the expert. We will take the main idea provided in [26] in the OCR field of application in order to be expanded and enhanced for the e-commerce domain where all those research aspects, related to context have been underestimated.

3 Problem Formulation

In the literature, context has had several alternative definitions in different fields and applications. The Webster's dictionary [29] defines context as "conditions or circumstances which affect some thing." In the data mining community, context is defined as those events which characterize the life of a customer and can determine a change in his/her preferences, status (e.g., prospect to actual), and affect the customer's value for a company [5]. In the context-aware systems literature, context was initially defined as the location of the user, the identity of people near the user, the objects around, and the changes in these elements [27]. In [4] a corpus of 150 definitions referring to different domains of cognitive sciences and related disciplines has been analyzed. In the field of e-commerce, the context has been defined as "the intent of" a purchase made by a customer, as supported by the anecdotal evidence provided at the beginning of this work. The same customer may buy from the same on line account different products for different reasons: a book for improving his/her personal work skills, a book as a gift for a partner, or an electronic device for his/her personal hobby. When the intent of the purchase varies, the user behavior is also supposed to change. As in the example, this kind of contextual information may be useful for building better user profiles and providing more accurate on-line recommendations. Given our definition of context, the problem can be formulated as follows.

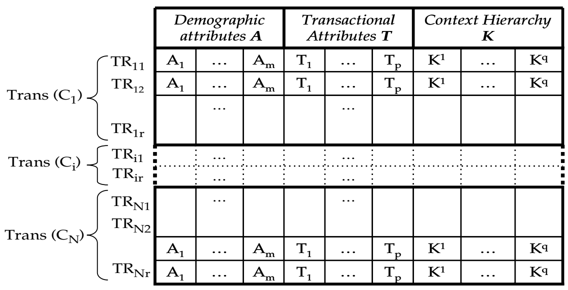


Fig. 1. User model-data structure

Let C be the customer base represented by N customers. Each customer C_i is defined by the set of m demographic attributes $A = \{A_1, A_2,\dots, A_m\}$, and a set of r transactions $Trans(C_i) = \{TR_{i1}, TR_{i2}, \dots,TR_{ir}\}$, where each transaction TR_{ij} performed by customer C_i is defined by a set of transactional attributes $T= \{T_1, T_2, \dots,T_p\}$. In addition, we also have a set of contextual attributes K associated with each transaction TR_{ir}. The table specifying all this demographic, transactional and contextual information about customers is presented in Figure 1. In general the set of contextual attributes K can have a complex structure reflecting the complex nature of this information. However, in this paper, we assume that domain K is defined by a set of q attributes $K^1,\dots, K^q$ having a hierarchical tree-shaped structure associated with it. In the tree structure the root represents the coarsest contextual knowledge while the leaves are the finest representation of the context, as in Figure 2. The values taken by attribute K^q define the finest (more granular) degree of contextual knowledge while K^1 the coarsest. For example, a customer C_i can be defined by the demographic attributes A= {*IDuser, Name, Age, Income*}, by the set of five transactions made by C_i, $Trans(C_i)$= {TR_1, TR_2, TR_3, TR_4, TR_5}, each transaction defined by the transactional attributes T= {*ProductID, StoreID, Price, TransactionTime*} and finally by a set of contextual attributes K describing the context ("the intent of") of each purchase.

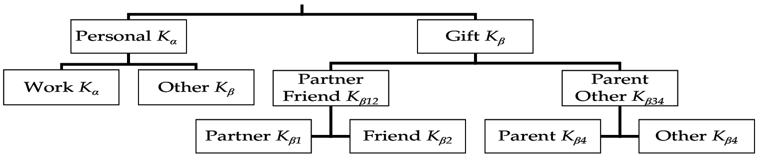


Fig. 2. Contextual hierarchy information for a purchasing transaction

Finally, the customer base C can be partitioned into several segments [15], [18] by computing h summary statistics S_i={S_{i1}, S_{i2}, …,S_{ih}} for customer C_i over the transactions made by that customer, each S_{ij} being defined as a statistics on some of the attribute in T across the transactions $Trans(C_i)$. Then customers can be clustered into segments in the space defined by these statistics. A model of customer behavior can be built in the following general form:

$$Y = f(X_1, X_2, ..., X_p)\ . \tag{1}$$

where X_1, X_2,..., X_p are some of the demographic attributes from A and some of the transactional attributes from T, and Y is the dependent variable to be predicted. Function f is a predictive function learned via different types of machine learning methods, such as logistic regression, decision trees or neural networks [30], that will be learned on the whole dataset shown in Figure 1. For instance, one may try to predict in which store the customer C_i will make a purchase, or which product will be bought, or the product's price. The predictive models of type (1) do not assume any contextual information since the contextual variable K is not a part of these models. Therefore, we call the models of this type *un-contextual.* In addition, we define *contextual* counterparts of predictive models (1), the model taking the following form:

$$Y = f_{K^q=\alpha}(X_1, X_2, \ldots, X_p)\ . \tag{2}$$

In model (2) only the transactions associated with the context K^q=a are used for building the model. For example, if the model is built for the computer science faculty from University X, where K^1=*"gift",* this means that only the *gift-related* transactions made by the CS faculty are used for building the model. Then the meaning of the expression "*context matters*" is that the contextual predictive models of type (2) *significantly* outperform in terms of predictive accuracy the un-contextual models of type (1) across different degrees of contextual knowledge [12]. The following two models are the formalization of the inferring process:

$$K^q = f(X_1, X_2, ..., X_p\ )\ . \tag{3a}$$

$$K^{Hier} = f(X_1, X_2, ..., X_p\ )\ . \tag{3b}$$

In model (3a) and (3b) the dependent variable, i.e. the variable to be predicted, is the contextual information. In Figure 3(a) and 3(b) the structures of the networks are presented for exploiting context represented as statistical dependencies; in each graph the structure is fixed and we only estimate its parameters. The use of graphical models for presenting contextual knowledge has been studied in [14], [26]. In model (3a) one contextual attribute per time is inferred. For instance, in the coarsest context degree each transaction is labeled with K^1, in the finest degree of contextual knowledge with K^3. In model (3b) the aim is to evaluate how the model can infer K^q using *the whole hierarchy* of context, rather than a single level of contextual knowledge (each transaction will be labeled with q contextual attributes $K^1, K^2,..., K^q$). For models (3a) and (3b) the predictive function f is defined as a Naïve Bayes (NB) network (Figure 3(a)) and a Bayesian Network (BN) in figure 3(b), respectively. In model (3b) hierarchy K^{Hier} consists of q variables ($K^1,..., K^q$) organized as shown in Figure 3(b), and where the last variable is K^q which is the same dependent variable as in the NB model (3a). Both BN and NB predict the same contextual variable K^q. We expect that by dropping the independence hypothesis of the NB model we should get better inference results for the model defined by (3b) in Figure 3(b). Another expectation is that the output of both models will be strictly influenced by the trade-offs between modeling error due to overly strict independence assumptions and estimation error of models that are too elaborate for the size of the available training set. Those expectations will be considered when the results of the comparison between the predictive performance of models (3a) and (3b) are analyzed.

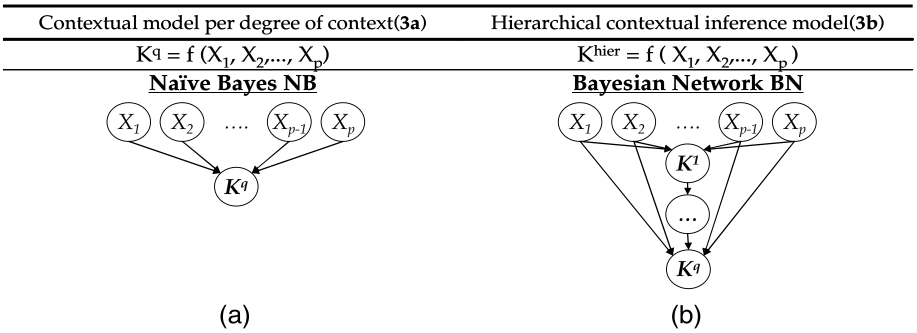


Fig. 3. Contextual Inference models

4 Experimental Setup

Since contextually rich datasets suitable for building personalized customer models are not readily available, as was explained in Section 1, we had to collect such data by ourselves in order to conduct our study. To this aim we developed a special-purpose browser to help users navigate a well-known e-commerce retail portal and purchase products on its site. This browser was made available to a group of student. Once a product was selected by a student in order to be purchased, the browser recorded the selected item, the purchasing price and other useful characteristics of the transaction.

In addition, the student had to specify the *context* ("intent of") in which the purchase was made. The data was pre-processed by excluding the students who made less than 40 transactions. The resulting number of students having at least 40 transactions was 556, and the total number of purchasing transactions for these students was 31,925. For each customer (student) we collected the following *demographic* data: age, previous studies, marital status, and composition of the family, place of living, hobbies and whether the student owned a car. The car ownership was used as a proxy for the income. The *transactional* data included item purchased, price, day, time, session duration, number of clicks per connection, and the time elapsed for each web page. The data set structure is described in Table 1.

Table 1. Data set structure

Demographic data A_i	*Type*	*Values/range*
1. Gender	Boolean	Male/Female
2. Age	Numerical	18-31
3. High School descr.	Nominal	Grammar, Professional, Private
4. Student description	Nominal	Outside, Traveling, Resident
5. Personal Car	Boolean	Yes/No
6. Hobby	Nominal	Reading, Dancing, Music, Electronics, Sports, Movies, Traveling, Informatics, Cooking, Cars, Arts, Photography, Collections, Fashion
Transactional data T_i	*Type*	*Values/range*
1. Visit Duration	Numerical	0-919 sec.
2. Price	Numerical	1-2000 $
3. N. of clicks	Numerical	1-35
4. Weekday	Boolean	Weekday/weekend
5. Store	Nominal	Electronics, home/garden, featured, Kid/baby, book/music, new
6.Purchase description	Boolean	Yes/No

The intent of purchase, i.e. the contextual information was collected at the beginning of each browsing session. The user was asked to specify whether the purchase would be intended for *personal* purposes or as a *gift*, *for which* specific personal purpose, and *for whom* the gift was intended. The overall contextual purchasing options were defined by the hierarchical structure presented in Figure 2. After collecting all the purchasing data for all the students, we built predictive models of their purchasing behavior for the contextual and un-contextual cases. Different experimental settings were obtained by varying the following parameters:

1. *Degree of contextual information*. The contextual models can be built by considering few values of K (rough knowledge of context) or as many as available (finer knowledge.)
2. *Granularity of customer segments*. A total of 4 unit of analysis have been used ranging from a single segment containing aggregated customer base to the 1-to-1 case when all the segments contain transactions of each single customer. Two intermediate levels included 100 segments and 10 segments. A predictive model is built for a cluster of customers.
3. *Types of predictive models*. We considered four different types of data mining classifiers modeling the function *f*, including decision trees and decision rules (JRip, J48, PART, NBTree) [30]. Those classifiers are trained and validated on the whole dataset.

4. *Dependent variables*. 3 transactional variables have been used for predicting customers' behavior (the positive or negative ending of each transaction, the day of the transaction, the store where the customer will purchase).
5. *Performance measures*. For providing statistical measurement of how well the classification functions correctly identifies or excludes a condition we used the predictive accuracy and the area under the ROC curve [30] as performance measures of predictive models.

For the contextual inference process we built predictive models of type (3a) and (3b) under different experimental settings. In the last case (hierarchical contextual model) the number of parameters is reduced: it has been used only one predictive function (BN and NB) in spite of four and the dependent variable is the context itself and not a transactional attribute. All the other settings are the same (degree of context, granularity of market, performance measures). The degree of contextual information is defined as follows.

The contextual information K is structured in a three-level hierarchy *(q=3)*, as shown in Figure 2, from a rough to a finer degree of knowledge. The contextual structures are deployed as follows. In the first level, the contextual variable K^1 takes two different values: K_α="*personal*" and K_β="*gift*". In the third and finer level K^3, the "*personal*" context is split in $K_{\alpha 1}$="*personal for work"* and $K_{\alpha 2}$=*"personal for other purposes",* the "gift" context is split in $K_{\beta 1}$="*gift for partner*", $K_{\beta 2}$="*gift for friends*", $K_{\beta 3}$="*gift for parents*" and $K_{\beta 4}$="*gift for others*". In the second level, four values are aggregated in two resulting in $K_{\beta 12}$=*"gift for partner and friends"* and $K_{\beta 34}$=*"gift for parent and other",* respectively.

5 Results

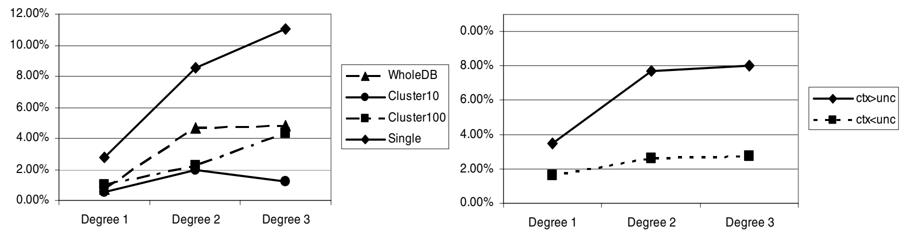
Given the number of experimental settings, resulting in a high number of contextual versus un-contextual comparisons, the more concise way to compare contextual model to the un-contextual one is computing the relative difference between the performance values as:

$$(\text{Performance}_{\text{con}} - \text{Performance}_{\text{unc}}) / \text{Performance}_{\text{unc}} . \quad (4)$$

where *con* refers to the contextual models and *unc* to the un-contextual model. A positive value means that the contextual model outperforms the un-contextual and viceversa. In both plots in Figure 4 are represented different degrees of contextual knowledge (from coarsest to finer) on the horizontal axis and on the vertical axis is measured the relative difference in performance averaged across all the experimental settings. In Figure 4(a) each of the four lines is plotted calculating performance formulation (4) where each line represents one level of customer segments. Figure 4(a) shows that the contextual models outperform the un-contextual for each type of customer segmentation. It also shows that in almost all cases of customer granularity, the curves representing the value of (4) are monotone, i.e. the finer the context degree the higher the value of (4). The graph also demonstrates that in most cases finer segmentation of the customer base leads to higher performance improvement when a contextual model is used instead of un-contextual. In fact, with the individual models of customers ("Single" in Fig. 4(a)) significantly outperform all other cases achieving

11% performance improvement for the finest granularity of contextual information ("Degree 3" in Fig. 4(a)).

Figure 4(b) presents the comparison between contextual and un-contextual models with respect to the degrees of contextual information but this time we have relaxed the market granularity assumption. In order to have a clearer representation, the positive values of (4) are computed separately from the negative values, and the absolute values are plotted in the graphs. The solid line plots the positive values of (4) per degree of context, while the dashed line plots the negative values of (4) in absolute terms. For the positive occurrence of (4) in Figure 4(b), the plot shows that the performance measure grows from the coarsest to the finest degree of contextual knowledge. In the negative occurrences of (4), there are stable and low performances.



(a) (b)

Fig. 4. Relative difference of performance per degree of context

Reporting the statistical significance of each comparison would have been impossible because of the large number of them (1152 in total). Figure 5 presents a summary of the statistical significance tests by reporting the percentage of comparisons with a statistical significance higher than 95%. A Wilcoxon test [3] was used for testing the null hypothesis (no difference between the two averages). In the graph the percentage of statistically significant comparisons are plotted against the degree of contextual information (on the x-axis) and for each customer granularity level (specified by different curves), for the cases in which the contextual models dominate the un-contextual.

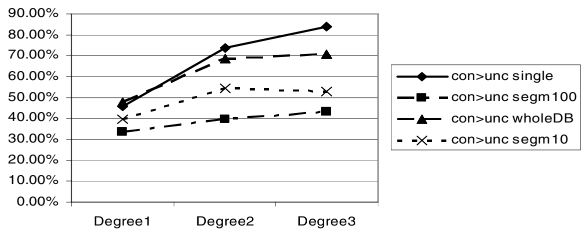


Fig. 5. Percentage of statistical significant comparisons

The values in Figure 5 are computed as follows: for each degree of context, the number of significant comparisons is divided by the overall number of comparisons and each line represents a different degree of customer granularity. For example, point 72% for Degree 2 for the "con>un-con single" line means that, fixing the experimental condition for the degree of context at second level and for customer granularity at single customer unit of analysis, the 72% of the possible comparisons where contextual outperform un-contextual are statistically significant. The number of statistically significant experiments where the contextual models outperform the uncontextual increases with the degree of knowledge of context and this is true per each customer granularity level (all curves are monotone). In particular, the best results are obtained when the single customer is the unit of analysis ("con>unc single" line in figure 5). The number of statistically significant experiments rises from 48% at the coarsest degree of contextual knowledge to 82% for the finest degree.

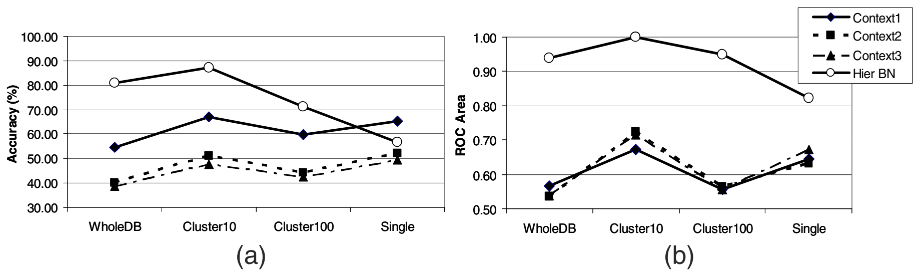


Fig. 6. Inference model performances per customer granularity

Figure 6 shows the results for the contextual inference problem. Both figures have been drawn applying the models (3a) and (3b). In Figure 6(a) performances is measured by the accuracy and in Figure 6(b) by the ROC area. The lines labeled as "context1", "context2" and "context3" are the plots of the inference performances achieved applying the model (3a) per each degree of context (K^1, K^2, K^3). The solid line with empty circles (specified by "Hier BN") represents the plot of the performance for model (3b) learned by the BN in Figure 3(b). Each line is plotted per degree of customer granularity (x-axis). In both figures the hierarchical model (3b) clearly outperforms each one of the models (3a), achieving high performance results, in particular the accuracy reaches the maximum value of 90% and the ROC area reaches the level of 100%. In both cases the maximum performance value was achieved when the unit of analysis is ten clusters of customers. Those results demonstrate that inferring the whole hierarchy of contextual knowledge in our experiment outperforms the approach in which each level per time is inferred. Moreover, trying to infer one level per time does not provide any improvement when the degree of contextual knowledge increases; for instance in the accuracy plot the inference of the coarsest degree of context it's higher than finer levels ("context1" outperforms "context2" and "context3"). Another interesting result is that the peak of performance achieved by the "Hier BN" at cluster 10 unit of analysis. This result shows that better inference performances can be achieved for a particular unit of analysis.

In terms of statistical significance we have evaluated two different aspects: the difference in performance between each of the four lines of the graph, and the difference in performance of each results of the "Hier BN" line. With the Friedman test (Non Parametric Repeated Measure Anova) [3] we have tested the null hypothesis that the performances of each of the four lines of the graphs are equal. For example, it means to evaluate how statistically significant is the difference in performance between the "context1" line and the "context2" and all possible combinations between all the shapes. The results are always statistically significant for the accuracy with p<0,001 at least. For the ROC area, the difference between the "Hier BN" and each of the single context lines is statistical significant (p<0,0001), while the difference in performance between each of the lines built with model (3a) are not. The results statistically support the statement that the hierarchical inference model (3b) outperforms (3a) and that inferring each degree of context alone is not useful. For evaluating the statistical significance of the difference in performance of each value point of the "Hier BN" model we have used the Kruskal-Wallis non parametric method [3]. The null hypothesis is that the performances obtained by (3b) are equal for each unit of analysis. For example it means to evaluate whether the 90% accuracy inference performance achieved when the unit of analysis is represented by ten clusters is statistically significant compared to the 80% value point achieved when the unit of analysis is the whole DB. The differences are always statistical significant with p<0,001 at least, supporting the statement that customer granularity indeed makes a difference in performing hierarchical inference models.

6 Discussions of Results

The results described in Section 5 present empirical evidence that the models built by taking into account rich contextual information usually provide better predictive performance in e-commerce applications. More specifically, the main conclusions of our study can be summarized as follows:

1. The degree of contextual information matters. The more we know about the context of a transaction, the better we can predict the customer's behavior.
2. Inferring the context is feasible. The predictive accuracy by which context is inferred may reach the value of 90%, and inferring the whole context hierarchy structure is better than inferring one hierarchical level per time.

Each point is discussed in detail below.

1. The degree of contextual information matters: the finer the knowledge about the context of a transaction, the better the predictive performance of a customer's behavior. As we move to finer degrees of contextual information, we observe higher values of performance gain. On average, as shown in Figure 4, knowing the finest degree of context leads to the highest gains in performance, growing from 3.5% (when *K* takes two values) to 8% (when *K* takes six values), in the cases in which the contextual models dominate the un-contextual. On the other hand, gathering finer degrees of contextual information can lead to a decrease in performance in those

settings in which the un-contextual model dominates the contextual. However, the loss is moderate (1.63% to 2.75%). The same evidence is provided by studying the statistical significance of the experimental results, as shown in Figure 5. Whatever the unit of analysis the number of statistically significant events grows when the degree of the contextual information grows and the contextual models dominate the un-contextual. The highest variations occur when the unit of analysis is the single customer. In fact, in this setting there is no case in which the un-contextual model outperforms the contextual models and the difference is statistically significant.
2. The inferring process is feasible. The results depicted in Figure 6 show that the inference of context from existing data is possible and inferring the whole context hierarchy structure is better than inferring one hierarchical level per time. In fact, predicting just the coarsest degree of context is on average 15% more accurate than the other two degrees. Inferring the whole hierarchy structure provides outstanding results, compared to the model (3a) but also in absolute terms. A 90% value of predictive accuracy and a value of 1 for the ROC area are definitely good enough for considering the opportunity of inferring the context off-line instead of getting the user involved in the process of gathering contextual information.

Another interesting result is related to the market granularity. The highest values of performances in inferring the context hierarchy is reached when the unit of analysis is a relatively low number of large customers segments (ten clusters). On the other hand, the highest value of accuracy in predicting customers' behavior (2) is reached when the unit of analysis is the single customer. This means that the hierarchical context inference model (3b) needs a large amount of data for inferring the context and this large amount of information can be provided in less granular market segments, while in model (2) the contextual effects get stronger when we build progressively smaller segments of customers, because providing contextual information, customer transactions pertaining to a particular context are reduced, making fewer data points to fit the model, while homogeneity of these transactions increases, making it easier to predict more accurately customer behavior in similar contexts.

Interpreting the inference results in the light of the literature on Bayesian probabilistic models leads to the following conclusions:

- The model (3b), relaxing the independence condition of NB models, remarkably well captures the inner dependencies between the attributes and the context without errors, as shown in Figure 6(a and b). In our case, the "expert supervision" in building the BN of model (3b) provides so efficient results that one should reject the opportunity of getting the customer involved in the process of gathering contextual information.
- Given the complexity of the network in (3b), the requested size of data for leveraging the trade off between the complexity of the network and the size of the sample needed for inferring the context is reached at some intermediate clustering level ("Cluster10" in Figure 6(a and b)).

Finally we can generalize that the inferring process is reliable only if the data analyst can properly select a good model, such as hierarchical BN and identifies proper segmentation level to make the best inference about the context.

References

1. Aaltonen, A., Huuskonen, P., Lehikoinen, J.: Context Awareness Perspectives for Mobile Personal Media. In: Aaltonen, Lehikoinen: Personal and Ubiquitous Computing (2005)
2. Adomavicius, G., Tuzhilin, A., Sankaranarayanan, R., Sen, S.: Incorporating Contextual Information in Recommender Systems Using a Multidimensional Approach. ACM Transactions on Information Systems 23(1), 103–145 (2005)
3. Barnes, J.W.: Statistical Analysis for Engineers and Scientists. McGraw-Hill, New York (1994)
4. Bazire, M., Brézillon, P.: Understanding Context Before Using It. In: Dey, A.K., Kokinov, B., Leake, D.B., Turner, R. (eds.) CONTEXT 2005. LNCS (LNAI), vol. 3554, Springer, Heidelberg (2005)
5. Berry, A., Linoff, G.: Data Mining Techniques: For Marketing, Sales, and Customer Support. Wiley Computer Publishing, New York (1997)
6. Bettman, J.R., Luce, M.F., Payne, J.W.: Consumer Decision Making: A Constructive Perspective. In: Tedeschi, M. (ed.) Consumer Behavior and Decision Making, pp. 1–42 (1991)
7. Bothorel, C., Chevalier, K.: How to Use Enriched Browsing Context to Personalize Web Site Access. In: Blackburn, P., Ghidini, C., Turner, R.M., Giunchiglia, F. (eds.) CONTEXT 2003. LNCS, vol. 2680, pp. 23–25. Springer, Heidelberg (2003)
8. Chen, G., Kotz, D.: A Survey of Context-Aware Mobile Computing Research. Department of Computer Science, Dartmouth College, Technical report (2000)
9. Dey, A.K., Abowd, G.D., Salber, D.: A Conceptual Framework and a Toolkit for Supporting the Rapid Prototyping of Context-Aware Applications. HCI 16(2), 97–166 (2001)
10. Flanagan, J.A.: Unsupervised clustering of context data and learning user requirements for a mobile device. In: Dey, A.K., Kokinov, B., Leake, D.B., Turner, R. (eds.) CONTEXT 2005. LNCS (LNAI), vol. 3554, Springer, Heidelberg (2005)
11. Fortu, O., Moldovan, D.: Identification of Textual Contexts. In: Dey, A.K., Kokinov, B., Leake, D.B., Turner, R. (eds.) CONTEXT 2005. LNCS (LNAI), vol. 3554, Springer, Heidelberg (2005)
12. Gorgoglione, M., Palmisano, C., Tuzhilin, A.: Personalization in Context: Does Context Matter When Building Personalized Customer Models? In: Perner, P. (ed.) ICDM 2006. LNCS (LNAI), vol. 4065, Springer, Heidelberg (2006)
13. Heath, T., Motta, E., Dzbor, M.: Uses of Contextual Information to Support Online Tasks. In: 1st AKT Doctoral Symposium, Milton Keynes, UK
14. Heckerman, D.: A tutorial on learning with Bayesian networks. Technical report, Microsoft Research, Redmond, Washington (1995)
15. Jiang, T., Tuzhilin, A.: Segmenting Customers from Population to Individuals: Does 1-to-1 Keep Your Customers Forever? IEEE TKDE 18(10), 1297–1311 (2006)
16. Kenny, D., Marshall, J.F.: Contextual marketing: the real business of the Internet, Harvard Business Review, 119–125 (2000)
17. Klein, N.M., Yadav, M.S.: Context Effects on Effort and Accuracy in Choice: an Enquiry into Adaptive Decision Making. Journal of Consumer Research 15(4) (1989)
18. Kotler, P.: Marketing Management. Prentice Hall, London (2003)
19. Leake David, B., Maguitman, A.G., Reichherzer, T.: Exploiting Rich Context: An Incremental Approach to Context-Based Web Search. In: Dey, A.K., Kokinov, B., Leake, D.B., Turner, R. (eds.) CONTEXT 2005. LNCS (LNAI), vol. 3554, pp. 254–267. Springer, Heidelberg (2005)

20. Lilien, G.L., Kotler, P., Moorthy, S.K.: Marketing Models. Prentice Hall, Englewood Cliffs (1992)
21. Luo, X.: The Performance Implications of Contextual Marketing for Business-to-Consumer Electronic Commerce: An Empirical Investigation. J. of Database Marketing 10(3) (2003)
22. Lussier, J.G., Olshavsky, R.W.: Task Complexity and Contingent Processing in Brand Choice. Journal of Consumer Research, 154–165 (1979)
23. Maamar, Z., Benslimane, D., Narendra, N.C.: What can context do for web services? Communications of the ACM 49(12), 98–103 (2006)
24. Pappas, A., Hailes, S., Giaffreda, R.: A design model for context-aware services based on primitive contexts. In: First International Workshop on Advanced Context Modelling, Reasoning And Management UbiComp (2004)
25. Prahalad, C.K.: Beyond CRM: C.K. Prahalad Predicts Customer Context is the next Big Thing. American Management Association MwWorld (2004)
26. Sarkar, P., Nagy, G., Veeramachaneni, S.: Modeling context as statistical dependence. In: Fifth International and Interdisciplinary Conference on Modeling and Using Context (2005)
27. Schilit, B., Theimer, M.: Disseminating active map information to mobile Hosts. IEEE Network 8(5), 22–32 (1994)
28. Schmidtke, H.R.: Granularity as a parameter of context. In: Dey, A.K., Kokinov, B., Leake, D.B., Turner, R. (eds.) CONTEXT 2005. LNCS (LNAI), vol. 3554, Springer, Heidelberg (2005)
29. Webster's new twentieth century dictionary of the English language (1980)
30. Witten, H., Frank, E.: Data Mining: Practical machine learning tools and techniques, 2nd edn. Morgan Kaufmann, San Francisco (2005)
31. Zhu, H., Madnick, S.E.: Structured Contexts with Lightweight Ontology. MIT Sloan Research Paper No. 4620-06 (2006)