Using Context to Improve Predictive Modeling of Customers in Personalization Applications

Cosimo Palmisano, *Member*, *IEEE*,
Alexander Tuzhilin, *Member*, *IEEE*, and
Michele Gorgoglione, *Member*, *IEEE*

Abstract—The idea that context is important when predicting customer behavior has been maintained by scholars in marketing and data mining. However, no systematic study measuring how much the contextual information really matters in building customer models in personalization applications has been done before. In this paper, we study how important the contextual information is when predicting customer behavior and how to use it when building customer models. It is done by conducting an empirical study across a wide range of experimental conditions. The experimental results show that context does matter when modeling the behavior of individual customers and that it is possible to infer the context from the existing data with reasonable accuracy in certain cases. It is also shown that significant performance improvements can be achieved if the context is "cleverly" modeled, as described in this paper. These findings have significant implications for data miners and marketers. They show that contextual information does matter in personalization and companies have different opportunities to both make context valuable for improving predictive performance of customers' behavior and decreasing the costs of gathering contextual information.

Index Terms—Personalization, context, data mining, user modeling, predictive modeling.

1 Introduction

The director of personalization of one of the major online retailing companies once received a nasty e-mail 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 lifestyle, which upset that customer. This true story is very symptomatic of problems pertaining to many personalization systems that often infer 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 [14], [24]. Getting such contextual data, such as the intent of a purchase, is not easy in many applications. For instance, it may not be practical to ask the customer in the previous example about the purpose of his purchase. Therefore, it is necessary to demonstrate that this contextual information indeed makes a significant difference in building better customer models in marketing and e-commerce applications. In this paper, we address the question of whether this additional contextual information matters, i.e., does it lead to building better personalized models of customer behavior, where by "better" we assume superior predictive performance. This problem is not trivial because it entails a tradeoff between transaction homogeneity and data sparsity: by providing contextual information, customer transactions pertaining to this 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. In data mining terms, this problem is related to the well-known bias-variance tradeoff, i.e., given contextual information, which effect dominates the other: decreased bias due to the homogeneity of transactions associated with the specified context or increased variance due to insufficient data associated with this context.

Since obtaining contextual information can be very hard, instead of acquiring this contextual information, one can try to infer it from the existing uncontextual data. Therefore, we also study the question of how easy it is to *infer* the contextual information and how accurately it can be done. Finally, once the contextual information is obtained, either through an acquisition or an inference process, it should be made useful for predicting customers' behavior and studying when these predictions are the most accurate. This is another topic addressed in this paper. The research questions that we have just described can be summarized as follows:

- C. Palmisano is with Aizoon Consulting S.r.l, Turin, Italy. E-mail: cosimo.palmisano@aizoon.it.
- A. Tuzhilin is with the Stern School of Business, New York University, 44 West 4th Street, Room 8-92, New York, NY 10012. E-mail: atuzhili@stern.nyu.edu.
- M. Gorgoglione is with the Dipartimento di Ingegneria Meccanica e Gestionale (DIMEG), Polytechnic of Bari, Viale Japigia 182, 70126 Bari, Italy. E-mail: m.gorgoglione@poliba.it.

Manuscript received 10 Sept. 2007; revised 28 Mar. 2008; accepted 18 Apr. 2008; published online 19 May 2008.
For information on obtaining reprints of this article, please send e-mail to: tkde@computer.org, and reference IEEECS Log Number TKDE-2007-09-0453.
Digital Object Identifier no. 10.1109/TKDE.2008.110.

1. Does context matter for building better models predicting customer behavior?
2. Is it necessary to acquire contextual information or it is possible to infer it from the data?
3. How do we exploit the inferred contextual information for modeling customer behavior?

In this paper, we answer these questions empirically by conducting an empirical study on two data sets across a wide range of experimental conditions. To answer the first question, we built two alternative customer models, one including contextual information and the other one not, and compared their predictive performances. To answer the second question, we built two Bayesian models (Naïve Bayes and Bayesian net) having the dependent variable specifying the contextual information and compared their performance in inferring the context. For the third question, we built three Bayesian "contextual" models (Naïve Bayes, Bayesian net, and Bayes net with latent contextual information) and compared them among themselves and with the uncontextual one to show that contextual information can be exploited for modeling customer behavior.

This study makes the following contributions to studying context in personalization applications. First, we demonstrate that context indeed matters when predicting customer behavior for individuals or small homogenous groups of customers and gets diluted during the process of aggregating customers' data. Second, granularity of the contextual information also matters: the more we know about the context of a transaction, the better we can describe the customer's behavior based on the context. Third, it is possible to infer fairly accurately the context for certain levels of customer segmentation using best-of-breed predictive models, such as certain types of Bayesian networks (BNs). Finally, if the context is not known from the external sources, it can first be inferred from the uncontextual data, as explained above, and then used for predicting customer's behavior. We show that the resulting model significantly outperforms the basic uncontextual model but underperforms the models having the explicitly known context.

2 Prior Work

The idea that the contextual information is important when predicting customer behavior is not new and much supporting evidence has been presented in the popular press. Context has become a major issue in several disciplines related to marketing and e-commerce. In e-commerce, the importance of context emerged in the market-space theory of business [27]. The theory maintains that value can be created for the customers in electronic markets by three elements: content, context, and infrastructure, where context refers to how firms offer information or content to the customers. The authors observed that the key to delivering successful e-commerce solutions has shifted from content to context. Further, according to Prahalad, "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*" [26].

New technologies, such as wireless Web, personal digital assistants, interactive television, and so forth, have created opportunities to target customers more effectively by providing real-time personalized information. This strategy is referred to as "*contextual marketing*" and is expected to result in numerous competitive advantages to e-business companies, such as loyal customers, cross selling and upselling opportunities, and so forth [21]. The research described in [21] empirically investigates the influence of contextual marketing on the perceived e-commerce outcomes of the online users and provides some initial evidence of positive influence.

The importance of including contextual information in recommendation systems has been demonstrated in [1], where the authors present a multidimensional approach that can provide recommendations based on additional contextual information besides the typical information on users and items used in most applications. However, Adomavicius et al. [1] do not investigate the use of context to build better predictive models of customers but focus only on product recommendations. Moreover, Adomavicius et al. [1] modeled context using the multidimensional method, whereas in this paper, it is modeled using contextual hierarchies, which constitutes a different method of modeling context. The use of interest scores assigned to topics has been applied to building contextual user profiles in recommender systems [37].

Marketing researchers have maintained that the purchasing process is contingent upon the context in which the transaction takes place since the same customer can adopt different decision strategies and prefer different products or brands depending on the context [6], [22]. According to Lilien et al. [20], "consumers vary in their decision-making rules because of the usage situation, the use of the good or service, and purchase situation." Therefore, accurate predictions of consumer preferences should depend on the degree to which we have incorporated the relevant contextual information. In the marketing literature, context has also been studied in the field of behavioral decision theory. In [22], context is defined as a task complexity in the brand choice strategy.

If the concepts of customer and 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 areas, such as context-aware systems, Web searching, and Web services. Context-aware systems are designed to exploit the contextual information available (e.g., where the user is and who is with her) to better serve the user and to adapt to the changes in the context [8], [12]. Contextual information has been proven to be helpful in information retrieval, although most existing systems base their retrieval decision solely on queries and document collections, whereas information about search context is often ignored [3]. The effectiveness of a proactive retrieval system depends on the ability to perform context-based retrieval, generating queries that return context-relevant results [19], [33]. In Web searching, context is considered as the set of topics potentially related to the search term. For instance, Lawrence [18] describes how contextual information can be used and proposes several specialized domain-specific context-based search engines. Integration of context into a Web services composition is suggested by Maamar et al. [23].

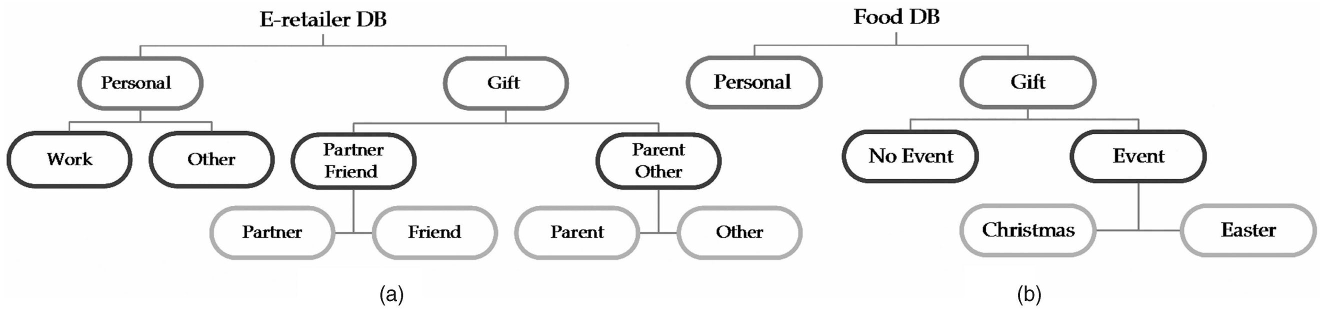


Fig. 1. Contextual information hierarchical structure: (a) e-retailer data set and (b) food data set.

A critical issue when modeling customer behavior is how to acquire the contextual information [11], [12]. In some circumstances, context is explicit, such as a person informing a company that she is moving to another city. Alternatively, the contextual information can also be inferred from the data or the environment, such as a change in location of the user detected by a mobile telephone company. There are two main approaches to inference: unsupervised and supervised learning. The former learns associations between context and user data without the explicit intervention of the user. The latter requires the user intervention at some point to label the context or define its certain characteristics. Usually, an expert defines ontologies [36], recommendation queries [2], or designs a context-specific model, such as a BN, in order to improve contextual inference [30].

Another issue related to this research is representation: context can be represented using various modeling methods including contextual graphs [9], conceptual maps [7], ontologies [15], [25], [34], [36], and hierarchical classifications [2], [32]. Also, the question of how to represent context, once it is inferred to model customers' behavior, has not been addressed before, except in [1] where it was only tangentially referred to and explored within recommender systems.

This paper is built on our previous work on contextual profiling [14], [24]. In particular, we have consolidated and enhanced our prior studies of questions 1 and 2 from [14] and [24], respectively. We also added question 3, expanded our data analysis to an additional data set, and integrated studies of questions 1 and 3 into a self-contained journal paper.

3 Problem Formulation

In this section, we first explain what we mean by "context," then how we model customer behavior, and finally, the methodology for comparing contextual and uncontextual models.

3.1 What Is Context

Many definitions of context can be found in the literature depending on the field of application, enabling technologies, and the available customer data. The Webster's dictionary defines context as "conditions or circumstances which affect some thing." In the data mining community, context is defined in [5] as those events that characterize the life of a customer and can determine a change in his/her preferences, status, and value for a company. Examples of context include a new job, the birth of a son, marriage, divorce, and retirement. 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 [31]. Other factors have been added to the previous definition. For instance, Brown et al. [10] includes the date, the season, and the temperature. Ryan et al. [29] add the physical and conceptual statuses of interest for a user. Dey et al. [12] include the user's emotional status and broaden the definition to any information that can characterize and is relevant to the interaction between a user and an application. Some associate the context with the user [12], [13], while others emphasize how context relates to the application [28], [35]. According to Prahalad [26], context has temporal (when to deliver), spatial (where), and technological (how) dimensions. Context is usually referred to the present situation, but sometimes the history of past is considered as well [12].

In this paper, context is defined as *the intent of* a purchase made by a customer in an e-commerce application. Different purchasing intents may lead to different types of behavior. For example, the same customer may buy from the same online account different products for different reasons: a book for improving her personal work skills, a book as a gift, or an electronic device for her hobby. In general, the context in which a customer performs a transaction is defined with a *set* of *contextual attributes* $\boldsymbol{K}$ that can have a complicated structure reflecting the complex nature of this information. Each contextual attribute K in $\boldsymbol{K}$ is defined by a set of q attributes $K = (K^1, \dots, K^q)$ having a *hierarchical structure*. The values taken by attribute K^q define *finer* (more *granular*) levels, while that of K^1 define *coarser* (less granular) levels of contextual knowledge.

For example, Fig. 1a presents a three-level hierarchy for the contextual attribute K specifying the intent of a purchasing transaction in an e-retailer application, considered in this paper. The root (coarsest level) of the hierarchy for K is defined by attribute $K^1 = \{Personal, Gift\}$, which labels each customer purchase as personal or as a gift. At the next, finer level of the hierarchy, "Personal" value of attribute K^1 is further split into a more detailed personal context: personal purchase made for the work-related or other purposes. Similarly, the *Gift* value for K^1 can be split into a gift for a partner or a friend and a gift for parents or others. Thus, the K^2 level is $K^2 = \{PersonalWork, PersonalOther, GiftPartner / FriendGift\ Parent / Other\}$. Finally, attribute K^2 can be split into level 3 of the hierarchy shown in Fig. 1a.

	Demographic attributes **A**			Transactional Attributes **T**			Context Hierarchy **K**		
	A_1	...	A_m	T_1	...	T_p	K^1	...	K^q
	...	...	...	...	...	...	...	...	...
Trans(C_i) TR_{j1}	A_{j1}	...	A_{jm}	$T_{j1,1}$	...	$T_{j1,p}$	...	...	...
TR_{j2}	A_{j1}	...	A_{jm}	$T_{j2,1}$	...	$T_{j2,p}$	...	...	...
TR_{jr}	A_{j1}	...	A_{jm}	$T_{jr,1}$	...	$T_{jr,p}$	...	...	...
	...	...	...	...	...	...	...	...	...

Fig. 2. Demographic, transactional, and contextual information about the customers and their transactions.

3.2 Customer Modeling

Let C be the customer base represented by N customers. Each customer C_i is defined by the set of m demographic attributes $A = \{A_1, A_2, \ldots, A_m\}$, and a set of r transactions $Trans(C_i) = \{TR_{i1}, TR_{i2}, \ldots, TR_{ir}\}$, where each transaction TR_{ij} performed by customer C_i is defined by a set of transactional attributes $T = \{T_1, T_2, \ldots, T_p\}$. In addition, we also have contextual information K associated with each transaction TR_{ir} of the form described in Section 3.1.

Fig. 2 presents a fragment of the customer table containing demographic, transactional, and contextual information about the customer C_i. For example, customer C_i can be defined by demographic attributes A $= \{IDuser,\ Name,\ Age,\ Income\}$, by five transactions $Trans(C_i) = \{TR_1, TR_2, TR_3, TR_4, TR_5\}$, each transaction defined by the transactional attributes:

$$T = \{ProductID,\ StoreID,\ Price,\ TransactionTime\},$$

and by the contextual attribute $K = (K^1, K^2, K^3)$ describing the context (e.g., "the intent") of each purchase,[1] as explained in Section 3.1 and shown in Fig. 1. Context hierarchy K in Fig. 2 specifies only one contextual attribute (note that $K^1, \ldots, K^q$ describe *levels* in this attribute). In general, however, we support multiple contextual attributes. Finally, the customer base C can be partitioned into several *segments* [21] by computing h summary statistics $S_i = \{S_{i1}, S_{i2}, \ldots, S_{ih}\}$ for customer C_i over the transactions $Trans(C_i) = \{TR_{i1}, TR_{i2}, \ldots, TR_{iki}\}$ made by that customer using various statistical aggregation and moment functions, such as mean, average, maximum, minimum, variance, and other statistical functions. For instance, for the transactions made by the customers in the previous example, the statistics can be $S_i = \{Average\ time\ spent\}$ or $S_{i+1} = \{Average\ price\}$. This means, among other things, that each customer C_i has a unique summary statistics vector S_i and that a customer is represented with a unique point in the space of these summary statistics. After generating such a data point per customer in the h-dimensional space of statistics $\{S_1, S_2, \ldots, S_h\}$, customers can be clustered into groups (segments) in that space using the clustering techniques described in [16]. Given segment p_α, of k customers $C_1, \ldots, C_k$, and their respective demographic $A_i = \{A_{i1}, A_{i2}, \ldots, A_{im}\}$ and transactional data $Trans(C_i) = \{TR_{i1}, TR_{i2}, \ldots, TR_{iki}\}$ for customers i in p_α, we want to build a single predictive model M_α on this segment of customers p_α:

1. We considered only one such contextual attribute to avoid clutter and to make Fig. 2 more manageable. As we pointed out in Section 3.1, there can be more than one contextual attribute in an application.

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

where dependent variable Y is one of the transactional attributes T_j, and independent variables $X_1, X_2, \ldots, X_p$ are all the transactional and demographic variables, except variable T_j, i.e., they form the set $T \cup A - T_j$. The performance of model M_α can be measured using some *fitness function* f mapping the data of this group of customers p_α into reals, i.e., $f(p_\alpha) \in \mathcal{R}$. For example, model M_α can be a decision tree built on data p_α of customers $C_1, \ldots, C_k$, for the purpose of predicting T_j variable "time of purchase" using all the transactional and demographic variables, except variable T_j as independent variables. The fitness function f of model M_α can be its predictive accuracy on the out-of-sample data or computed using 10-fold cross validation. 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 *uncontextual*. We define *contextual* counterparts of predictive models (1), where the model takes the following form:

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

$$Y = f\ \left(X_1, X_2, \ldots, X_p, K^q\right), \tag{2b}$$

where the two models (2a) and (2b) constitute two different ways of creating a contextual model. Model (2a) indicates that only transactions associated with a particular value of the context attribute $K^q = \alpha$ are used for building the model. In this case, the contextual information is used as a label for filtering customer transactions and then dropped. For example, if model (2a) 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. In model (2b), the generic contextual attribute K^q is considered as an independent variable, such as the demographic and transactional attributes $X_1, X_2, \ldots, X_p$. This means that it is used as one of the attributes for predicting Y. Because of the hierarchical structure of the set of attributes K, the number of contextual models can vary depending on how fine the contextual knowledge is, i.e., at which level of the hierarchy value q is defined in K.

One interesting question when building contextual models is where to place purchasing transactions of customer C_1 when she bought a gift for customer C_2: should such a transaction be associated with the purchasing history of customer C_1 or C_2? In this paper, we associate such purchases with customer C_1 and not C_2 for the following reasons. First, these purchases reflect perceptions of customer C_1 about what customer C_2 needs, not the real traits and needs of customer C_2. Second, even though the user may want to interpret expectations and preferences of another individual, it would be very unusual to model behavior of a person by observing the behavior of another individual. Third, when building a model for customer C_1, the demographical and transactional data used in this behavioral model are those related to customer C_1. One way to handle this problem of gifts is to define an appropriate

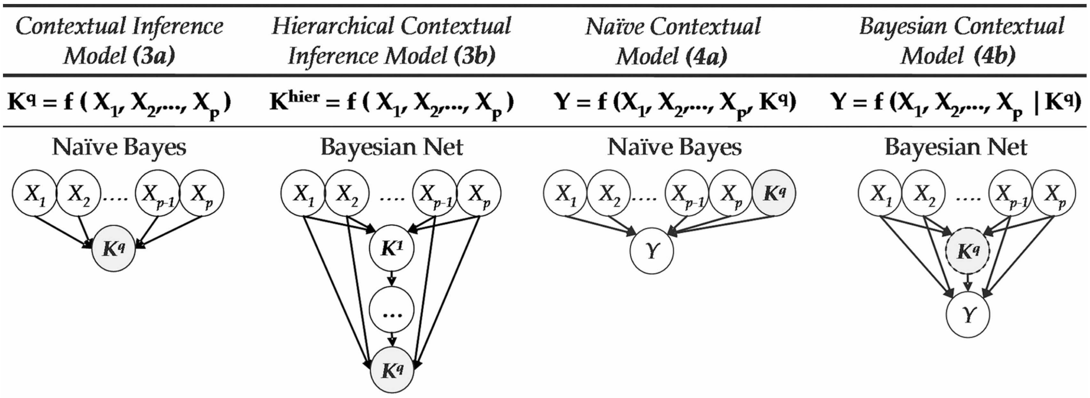


Fig. 3. Graphical representation of the predictive models.

context of gifts and purchases for others and treat such purchasing behavior in these contexts. This provides for extra flexibility because we can treat such purchasing transactions differently in different contexts.

For answering the first research question (*Does context matter?*), a comparison between performance results of predictive models is performed for the uncontextual (1) and the contextual (2a) models across a wide range of experimental conditions.

The inference problem (*Is it feasible to infer context?*) can be specified by replacing the dependent variable Y in (1) with the context variable K as

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

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

In models (3a) and (3b), the dependent variable is the contextual information. In (3a), one contextual attribute is inferred at a time. For instance, for the coarsest degree of contextual knowledge, each transaction is labeled with K^1, and for the finest degree with K^3. In model (3b), the aim is to evaluate how the model can infer K^q given the *whole hierarchy* of context, rather than a single level of contextual knowledge (each transaction is labeled with q contextual attributes $K^1, K^2, \ldots, K^q$). For model (3a), f is a predictive function learned via different types of machine learning methods while for model (3b) the predictive function f is defined as a BN [30]. We have studied whether inferring K^q, given the whole hierarchy, is better than inferring one degree at a time.

For the third question (*How do we exploit context?*), we used the following *contextual* models:

$$Y = f\,\big(X_1, X_2, \ldots, X_p, K^q\big), \tag{4a}$$

$$Y = f\,\big(X_1, X_2, \ldots, X_p | K^q\big). \tag{4b}$$

Given the set of customer transactions represented by demographical, transactional, and contextual attributes, the aim is to predict the behavior of the customer. In model (4a), the contextual attribute K^q is used as an independent variable together with attributes $X_1, X_2, \ldots, X_p$. In (4b), K^q is modeled as a variable statistically dependent on variables $X_1, X_2, \ldots, X_p$. We used the Naïve Bayes method to model function f in (4a) and a BN to model f in (4b).

In Section 5, we compare performances of the uncontextual model (1), the two contextual models (4a) and (4b), and the following latent model:

$$Y = f_{K^q}\,(X_1, X_2, \ldots, X_p). \tag{4c}$$

In Fig. 3, the structures of models (3a) and (3b) and (4a)-(4c) are presented as directed acyclic graphs [30]; each graph describes the dependencies of the attributes for each specific predictive model. Models (4b) and (4c) are defined with the same BN in Fig. 3 (rightmost BN); but model (4b) uses the given contextual information while model (4c) infers the unknown contextual information.

4 Experimental Setup

To answer the research questions 1-3 stated in Section 1, we have conducted experiments across the following different experimental conditions:

1. data sets,
2. degree of contextual information,
3. granularity of customer segments,
4. types of predictive models,
5. types of dependent variables used in these models, and
6. types of performance measures.

We describe each of these experimental conditions in detail in the rest of this section.

4.1 Data Sets

The experiments have been conducted using two customer data sets: a gourmet food data set and an e-retailer data set. We describe each data set below.

E-retailer data set. Since contextually rich "industrial-strength" data required in our studies was not readily available, we decided to collect such data on our own as follows: First, a *special-purpose browser* was developed 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 students. While navigation was real, purchasing was simulated (no real money was spent). The user could visit any page of the portal and use all the browsing and navigational activities except the actual purchase function. Once a product was selected to be purchased, the

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

(a)

Demographic data Ai	Values/range
1. Gender	Male/Female
2. Age	18-70
3. Billing region	20 regions
4. Newsletter	Yes/No
5. Job Position	Free-lance, unemployed, employee, housewife, student, retired
Transactional data T_i	**Values/range**
1. Region delivery	20 regions
2. Year	2004/2005
3. Day delivery	Monday to Sunday, np
4. Time delivery	Morning,Afternoon, np
5. Time Purchase	Morning, afternoon, Evening, Night
6. Id_Product	1,2,3,4
7. Quantity	1 to 6
8. Discount	Yes/No

(b)

Fig. 4. Attributes of the data sets used in our study: (a) e-retailer attributes and (b) gourmet food attributes.[3]

browser recorded the selected item, the purchasing price, and other useful characteristics of the transaction. The browser was directly linked to a data set, where all the customer information about browsing and purchasing activities on the portal was automatically recorded. The *contextual information* (*intent of* the purchase) 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 transactional data were collected by the file logs and by using clickstream techniques. No restrictions were imposed on the participants either in terms of the products they could purchase or the amount of money to spend. They were only recommended to buy on the order of 50 items. The data were preprocessed by excluding the students who made less than 40 transactions and eliminating the students who had any kind of misleading or abnormal behavior such as buying 40 products in a short time range or buying 40 times the same product. In order to discourage misleading behavior, a reward system was established that encouraged "normal and reasonable" purchasing activities. The resulting number of students having at least 40 transactions and showing "normal" browsing behavior was 556, and the total number of purchasing transactions for these students was 31,925. For each customer (student), the following demographic data were collected: age, previous studies, marital status, composition of the family, place of living, hobbies, and whether the student owned a car. The transactional data included item purchased, price, day, time, session duration, number of clicks per connection, and the time elapsed for the web page. Fig. 4a reports the set of attributes $X_1, X_2, \ldots, X_p$ used for building predictive models. Table 1a provides the average number of transactions for e-retailer data set across different levels of customer segmentation and degree of contextual information.[2]

Gourmet food data set. The second data set used in the experiments is a data set of customer purchases of gourmet food collected online by a European food distributor. The original data set included purchases made by customers in 2004 and 2005, with 7,800 transactions performed by 250 customers (only few customers bought the food online; most still preferred real stores). The data set schema of demographic and transactional attributes is presented in Fig. 4b. Note that contextual attributes are not included in Fig. 4b because the contextual information was not always available but often hidden in the data. The products available online are purchased for both making gifts in special events and for the personal use by customers located in regions not covered by the distribution channels. In few cases, the intent of making a gift was explicitly expressed in a note attached by the customer at the end of the transaction. Considering the small size of the data set and the lack of explicit contextual information, the gourmet food data set was preprocessed in two steps: 1) labeling each transaction with a reliable context and 2) artificial enhancement of the original data set to increase its size. In the first step, a triple check identification process was applied: the mismatch between the billing and the mailing address, the time window in which the transaction was performed (Christmas, Easter, and other special events and holidays), and the presence of a greeting note (the data set contained a special field where the customers could place their instructions pertaining to the delivery, greeting notes, and so forth). The combination of those three elements has been used for selecting the contextual information. For example, when the billing address of a transaction was in a region different from the delivery address, the transaction was labeled as "gift." If the time window was in a special event period and/or the customer noted explicitly that the purchase was for a specific kind of gift, the gift transaction was labeled with finer degree of contextual knowledge. This check process was used to manually label data with contextual information. In order to make the contextual knowledge finer and building a hierarchical structure, the "gift" transactions were associated with a particular event. The following labels were used:

2. "Whole DB" specifies transactions for the whole customer base and "Single" specifies the average number for a single customer. Other columns in Table 1a, such as Cluster10 and Cluster100, will be explained below.

3. The Id_Product transactional attribute refers to the four different products sold in the gourmet food data set.

TABLE 1
Average Number of Transactions for (a) E-Retailer and (b) Food Data Sets for Levels of Context and Customer Segmentation

e-retail	No. of transactions			
	Whole DB	Cluster10	Cluster100	Single
Degree1	15963	1596	160	29
Degree2	7981	798	80	14
Degree3	5321	532	53	10

(a)

e-food	No. of transactions			
	Whole DB	Cluster10	Cluster100	Single
Degree1	39535	3954	395	107
Degree2	26357	2636	264	72
Degree3	19768	1977	198	54

(b)

1. *Christmas:* gift transactions made between 1 December and 6 January (37 days per year).
2. *Easter:* gift transactions made Three weeks before the Easter day (21 days per year).
3. *No_event:* gift transactions made in other periods of the year.

The hierarchical contextual structure for this data set is represented in Fig. 1b.

The data set was enlarged by 1) adding new artificially generated transactions per customer that preserve the distributions of all the transactional attributes based on the actual "real-world" transactions and 2) augmenting the number of customers by using a set of customers registered for the newsletter service (we used demographic information of the "transactionless" customers from the newsletter, even if they did not make any online purchases). The customers' transactional histories were built by clustering the active customers using demographic profiles and then matching the profiles of "transactionless" customers with these clusters. Further, a 10-time resampling with 30 percent degree of noise in each transactional attribute was used.[4] The final size was 368 customers, 79,070 transactions, and 214 transactions per customer on average. Table 1b reports average number of transactions for different customer segmentation levels.

4.2 Degree of Contextual Information

The contextual information K^q is structured in a three-level hierarchy in both data sets from a coarse ($K_\alpha =$ "$personal$", $K_\beta =$ "$gift$") to a finer degree of knowledge. Fig. 5 summarizes the value of K for each level of context in each data set, based on Figs. 1a and 1b.

4.3 Granularity of Customer Segments

The behavioral models can be built for different groups of customers (unit of analysis) depending on how finely the customer base is partitioned. As one progresses from the aggregate to the individual level of granularity, the groups of customers become increasingly more homogenous, assuming good clustering methods are used, thus making predictions potentially more accurate [16]. However, individual models run into the data sparsity problem. Thus, experimental analysis is crucial for studying under which conditions the knowledge of context dominates the sparsity effect. The following levels of analysis were applied in the experiments:

1. *Whole customer base.* One model is built to predict the behavior of the whole customer base.
2. *Cluster10.* One predictive model is built for each one of 10 macrosegments. The segments were generated by applying the Farthest First clustering method[5] to the customer base, where each customer is defined by a vector of summary statistics attributes.
3. *Cluster100.* The units of analysis are 100 microsegments. Both the summary statistics and the clustering algorithms are the same as in the Cluster10 case.
4. *Single customer.* A predictive model of customer behavior is applied to the transactions made by each single customer.

4.4 Types of Predictive Algorithms

For the first and second research questions (*Does context matter? Can we infer it?*), four types of Weka classifiers were used for building predictive models: Naïve Bayes, J48 (a version of a decision tree), PART (an association rule classifier), and JRIP (a rule-based classifier) [4]. They were selected because they are both popular classification methods and computationally fast: computation time is crucial because 220,704 models were generated in our experiments.

For the third research question (*How do we exploit context?*), probabilistic classification algorithms were used, in particular Naïve Bayes and BNs. A BN algorithm was also used to evaluate the performance of the hierarchical context inference model. The performance of the algorithms was carried out by using the 10-fold cross-validation method.

4.5 Dependent Variables

The following dependent variables Y have been chosen in our experiments. In the e-retailer case, the model predicts 1) whether a customer or a group of customers will make a *purchase* or not, 2) the *day of the week* a customer will perform a transaction, regardless of whether the session ends with a purchase or not, and 3) in which *store* a transaction will be made.

In the gourmet food data set, the model predicts 1) which of the four different types of *products* will be purchased by the customer, 2) the total *number of items* in the shopping cart, and 3) if the customer will perform the transaction with a *discount*.

For the second research question (*Can we infer context?*), this experimental condition was not applied because the dependent variable is the context attribute itself.

4.6 Performance Measures

We used two performance measures in our experiments. The first is the predictive accuracy [4] computed as the ratio between the number of correctly classified instances and the total number of classified instances. The second is the area under the ROC curve (AUC), which takes into account class distribution and the cost of errors [4]. Further, if *Cluster100*

4. We have selected 30 percent degree of noise because it provided the best compromise between preserving the distribution of transactional attributes and reducing the number of identical transactions. We selected the 10-time resampling ratio because we wanted to have at least 10 transactions per customer for each context, and this was the minimally sufficient number for us to achieve this goal.

5. We selected the Farthest First clustering algorithm because, as shown in [16], it provided better performance for similar types of problems compared to other popular clustering methods, including the K-means.

Contextual level	E-Retailer DB	Food DB
K^1 (rough)	Personal, Gift	Personal, Gift
K^2	PersonalWork, PersonalOther, GiftPartner/Friend, Gift Parent/Other	Personal, GiftNoEvent, GiftEvent
K^3 (finest)	PersonalWork, PersonalOther, GiftPartner, GiftFriend,GiftParent, GiftOther	GiftNoEvent, GiftEaster, GiftChristmas

Fig. 5. Values taken by each contextual attribute per each DB.

is the unit of analysis, the experiment would end up with 100 predictive accuracy measures and 100 AUC measures, for each experimental setting. Since the aim of the research is to compare the performance of the uncontextual and contextual models, it is necessary to compare the *distributions* of performance measures obtained by applying the models in each experimental condition and see if there are statistically significant differences between them. For each of the three research questions, a different type of statistical test has been applied and a different null hypothesis has been set up. These tests will be described in Section 5.

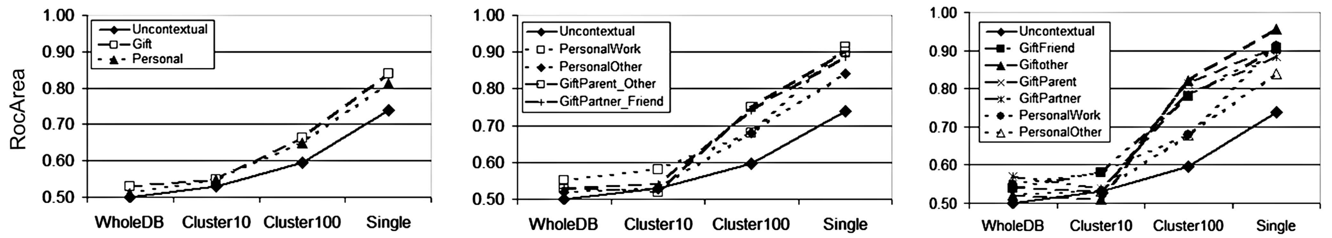
5 Results

5.1 Does Context Matter?

The aim of this study is to experimentally demonstrate that customer behavior changes when the intent of purchase changes by a comparative analysis of context-dependent and context-independent customer models in different experimental settings. Given the number of experimental settings (three degrees of contextual information, four customer granularities, two data sets, four classifiers, three dependent variables, two performance measures), the total number of generated models was more than 182,000 and the number of tables reporting the performance comparison results is 480, which constitutes a challenging problem to present. To give a "flavor" of the results, Fig. 6 presents three graphs generated by plotting the values of the AUC measure for the JRIP classifier and dependent variable Purchase for the e-retailer data set for different degrees of contextual information. The graphs are presented in the order of progressively more refined contextual information, from Fig. 6a corresponding to the coarsest level (K^1) to Fig. 6c corresponding to the finest (K^3), and different customer granularities on the x-axis. The points for *Cluster10* and *Cluster100* represent the performance averages taken over segment distributions. The graphs show that for these particular experimental settings the predictive performance improves from the aggregate to the single case, except in one case ($K^3 =$ "gift to the partner", third graph, second level of granularity). Moreover, all the contextual models show a better predictive performance compared to the uncontextual, except for two cases (*Cluster10* in Fig. 6c) where the difference is very small. The performance differences between uncontextual and contextual models also increase for finer degrees of contextual information. All these differences are statistically significant ($p < 0.01$), except for the *PersonalOther* context at *Cluster10* in Fig. 6b. Similar charts can be plotted for all other 479 cases mentioned above. It turns out that the same type of behavior has been observed for these charts as in Fig. 6 across other classifiers (J48, PART, Naïve Bayes), dependent variables (weekday, store), and performance measures (predictive accuracy). Although for these other charts the curves are not always monotonic, the predictive performance of the contextual models is usually higher than that of the uncontextual model for the finest level of customer granularity and the degree of contextual information. Instead of plotting individual graphs, a more concise representation can be obtained by computing the average values of performance for the two main experimental settings: degree of context and customer granularity. The averages are taken over the four predictive models, the three dependent variables, and the two performance measure distributions. However, the two performance measures vary in the ranges, namely, [0, 1] and [0.5, 1] for predictive accuracy and AUC, respectively. A reasonable way to compare each contextual model (2a) to the uncontextual (1) is by computing their relative performance differences:

$$\mathrm{Diff} = (\mathrm{Perf}_{\mathrm{con}} - \mathrm{Perf}_{\mathrm{unc}})/\mathrm{Perf}_{\mathrm{unc}}, \tag{5}$$

taken as averages over some experimental settings, where con refers to the contextual model and unc to the uncontextual model. A positive value of $Diff$ means that the contextual model outperforms the uncontextual, and the negative—otherwise. Tables 2a and 2b report the values of (5) computed as the average over four classifiers and two performance measures for each data set. Since there are many more pluses than minuses, Tables 2a and 2b demonstrate that the contextual models outperform the uncontextual in most



(a) (b) (c)

Fig. 6. Example of performance (ROC area) comparisons of JRIP predictive models for different levels of contextual information for the e-retailer data set: (a) first degree of contextual knowledge, (b) second degree, and (c) third degree.

TABLE 2
Average Performance for the (a) E-Retailer and (b) Food Data Sets

			WholeDB	*Cluster10*	*Cluster100*	*Single*
E-retailer Dataset	Dep.Variable: WeekDay	Gift-Uncontextual	-0,02%	-0,67%	-1,23%	**1,56%**
		Personal-Uncontextual	-0,02%	-0,52%	**1,03%**	**2,96%**
		GiftParent_Other-Uncontextual	**1,99%**	**0,66%**	-2,88%	**5,37%**
		GiftPartner_Friend-Uncontextual	**3,00%**	**4,55%**	-1,35%	**2,64%**
		PersonalWork-Uncontextual	**4,37%**	**4,44%**	-3,02%	**8,24%**
		PersonalOther-Uncontextual	**2,79%**	**2,56%**	**1,37%**	**5,24%**
		GiftFriend-Uncontextual	-0,26%	**0,71%**	-1,71%	**7,41%**
		Giftother-Uncontextual	**5,80%**	**2,41%**	**1,37%**	**18,07%**
		GiftParent-Uncontextual	**1,24%**	-0,37%	-1,64%	**6,18%**
		GiftPartner-Uncontextual	**5,99%**	**5,41%**	**2,00%**	**9,68%**
		PersonalWork-Uncontextual	**4,37%**	**4,44%**	-3,02%	**8,24%**
		PersonalOther-Uncontextual	**2,79%**	**2,56%**	**1,37%**	**5,24%**
	Dep.Variable: Purchase	Gift-Uncontextual	**0,76%**	**1,77%**	**3,99%**	**6,27%**
		Personal-Uncontextual	**0,25%**	**0,38%**	**3,44%**	**5,02%**
		GiftParent_Other-Uncontextual	**1,15%**	**0,08%**	**11,29%**	**10,95%**
		GiftPartner_Friend-Uncontextual	**1,60%**	**1,83%**	**10,20%**	**9,79%**
		PersonalWork-Uncontextual	**3,30%**	**3,40%**	**5,54%**	**11,39%**
		PersonalOther-Uncontextual	**0,05%**	**0,04%**	**5,74%**	**6,60%**
		GiftFriend-Uncontextual	**2,24%**	**2,65%**	**12,83%**	**10,71%**
		Giftother-Uncontextual	**2,35%**	-0,92%	**16,59%**	**15,15%**
		GiftParent-Uncontextual	**1,56%**	**2,16%**	**15,99%**	**11,65%**
		GiftPartner-Uncontextual	**4,03%**	**2,30%**	**13,87%**	**9,77%**
		PersonalWork-Uncontextual	**3,30%**	**3,40%**	**5,54%**	**11,39%**
		PersonalOther-Uncontextual	**0,05%**	**0,04%**	**5,74%**	**6,60%**
	Dep.Variable: Store	Gift-Uncontextual	-2,98%	**0,92%**	-4,76%	-3,42%
		Personal-Uncontextual	**6,30%**	**1,35%**	**3,61%**	**4,25%**
		GiftParent_Other-Uncontextual	-6,94%	-7,40%	-8,30%	**3,75%**
		GiftPartner_Friend-Uncontextual	**3,44%**	**1,94%**	**0,04%**	**6,43%**
		PersonalWork-Uncontextual	**31,11%**	**11,05%**	**3,96%**	**25,91%**
		PersonalOther-Uncontextual	**10,26%**	**0,82%**	**4,57%**	**6,35%**
		GiftFriend-Uncontextual	4,11%	-4,66%	**2,62%**	**9,66%**
		Giftother-Uncontextual	-0,83%	-4,46%	**0,39%**	**15,57%**
		GiftParent-Uncontextual	-5,39%	-10,06%	-6,56%	**5,33%**
		GiftPartner-Uncontextual	**14,41%**	**3,91%**	**3,96%**	**15,53%**
		PersonalWork-Uncontextual	**31,11%**	**11,05%**	**3,96%**	**25,91%**
		PersonalOther-Uncontextual	**10,26%**	**0,82%**	**4,57%**	**6,35%**

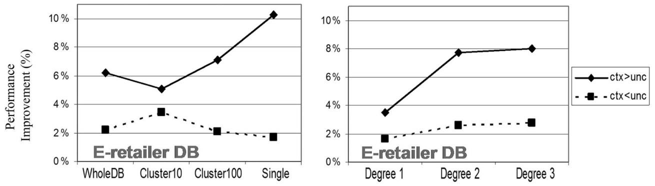
(a)

			WholeDB	*Cluster10*	*Cluster100*	*Single*
Food Dataset	Dep.Var.: Discount	Gift-Uncontextual	**4,74%**	**5,80%**	**6,01%**	**5,28%**
		Personal-Uncontextual	**3,07%**	-3,90%	-1,29%	-1,54%
		Personal-Uncontextual	**3,07%**	-3,90%	-1,29%	-1,54%
		GiftEvent-Uncontextual	**3,07%**	**5,37%**	**6,05%**	**6,43%**
		GiftNoEvent-Uncontextual	**0,47%**	-3,98%	**1,94%**	-0,04%
		GiftNoEvent-Uncontextual	**0,47%**	-3,98%	**1,94%**	-0,04%
		GiftChristmas-Uncontextual	**3,96%**	**3,76%**	**4,04%**	**6,06%**
		GiftEaster-Uncontextual	**3,85%**	-4,19%	**4,87%**	**5,36%**
	Dep.Variable:Product	Gift-Uncontextual	**3,03%**	**7,38%**	**3,39%**	**4,77%**
		Personal-Uncontextual	-0,90%	-1,69%	-1,04%	**1,25%**
		Personal-Uncontextual	-0,90%	-1,69%	-1,04%	**1,25%**
		GiftEvent-Uncontextual	**8,81%**	**6,47%**	**2,99%**	**4,45%**
		GiftNoEvent-Uncontextual	**7,85%**	**1,87%**	**1,66%**	**2,13%**
		GiftNoEvent-Uncontextual	**7,85%**	**1,87%**	**1,66%**	**2,13%**
		GiftChristmas-Uncontextual	**7,96%**	**6,15%**	**2,62%**	**4,18%**
		GiftEaster-Uncontextual	**6,65%**	-1,58%	**3,49%**	**4,88%**
	Dep.variable:Quantity	Gift-Uncontextual	-0,07%	**4,26%**	**2,14%**	**1,73%**
		Personal-Uncontextual	**1,30%**	**0,31%**	-0,26%	**0,67%**
		Personal-Uncontextual	**1,30%**	**0,31%**	-0,26%	**0,67%**
		GiftEvent-Uncontextual	**1,30%**	**5,81%**	**2,53%**	**2,25%**
		GiftNoEvent-Uncontextual	-1,99%	**2,42%**	**1,75%**	**2,56%**
		GiftNoEvent-Uncontextual	-1,99%	**2,42%**	**1,75%**	**2,56%**
		GiftChristmas-Uncontextual	-0,99%	**6,90%**	**3,03%**	**2,09%**
		GiftEaster-Uncontextual	-3,07%	**3,33%**	**3,29%**	**3,23%**

(b)

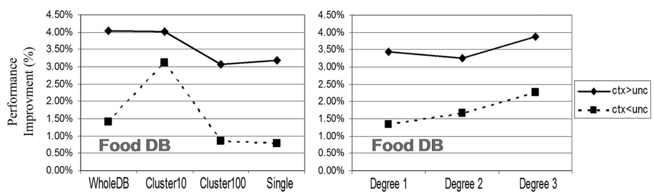
of the experimental settings. The crucial results from Table 2 are summarized in Figs. 7a and 7b, where positive values of $Diff$ are computed separately from the negative values and only the absolute values are plotted, thus differentiating between the cases in which the contextual models outperform the uncontextual ones $(\mathrm{ctx} > \mathrm{unc})$ and vice versa $(\mathrm{ctx} < \mathrm{unc})$. In Figs. 7a and 7b, the results are averaged over the four predictive models, the three dependent variables, and the two performance measures. In Fig. 7a, the results are represented per customer granularity while in Fig. 7b per degree of context knowledge. In particular, Fig. 7a demonstrates that in most cases a finer segmentation of the customer base leads to higher performance when a contextual model is used instead of uncontextual. In fact, the individual models of customers significantly outperform all other cases, achieving 11 percent performance improvement for the finest degree of contextual information. The small decrease in performance when the unit of analysis moves from the whole set of customers (*WholeDB*) to 10 segments of customers (*Cluster10*) is due to the fact that the segments are still quite large and not homogenous enough to assure better predictive performance. The homogeneity starts leveraging the effect of data sparsity when clusters are smaller and more homogeneous, as in *Cluster100*. Fig. 7b presents the comparison between contextual and uncontextual models with respect to the degrees of contextual information, but this time the customer granularity assumption has been relaxed. Fig. 7b shows that the finer the context, the more it matters. Moreover, this is also true across most of the cases of customer granularity. Results similar to the e-retailer data set were obtained for the gourmet food data set. We present in Figs. 8a and 8b the outcomes of the analyses similar to Fig. 7.

Since the comparisons entail comparing the averages of two distributions of model performance results across different clusters, a test of statistical significance needs to be performed. The null hypothesis is that there is no difference between the two averages, i.e., the performance of the uncontextual model is equal to that of the contextual models. Because the expectation is that context matters, the test is directional [4]. Since the two models are built for the same unit of analysis (single customer, segment, whole set



(a) (b)

Fig. 7. E-retailer DB: improvement in performance per (a) customer granularity and (b) degree of context.



(a) (b)

Fig. 8. Food DB: improvement in performance per (a) customer granularity and (b) degree of context.

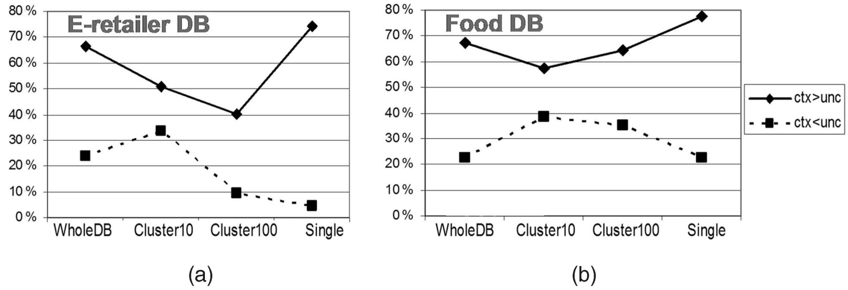


Fig. 9. Statistical significance per customer granularity: (a) e-retailer data set and (b) food data set.

of customers), the two samples are related. The distribution is not normal and a Wilcoxon test [4] was used for testing the null hypothesis. Since reporting the statistical significance of each comparison would be impossible because of the large number of them (1,152 only for the e-retailer data set), Fig. 9 presents a summary of these comparisons by reporting the percentage of comparisons with a statistical significance higher than 95 percent. The cases in which the contextual models significantly outperform the uncontextual (solid line) are plotted separately from the cases where the uncontextual dominate the contextual models (dashed lines). The values in Figs. 9a and 9b, for the e-retailer and gourmet food cases, respectively, are computed as follows: for each degree of customer granularity, the number of significant comparisons is divided by the overall number of comparisons.

The results related to the first research question can be summarized as follows.

5.1.1 Knowing the Context Matters for the Personalized (One-to-One) Models

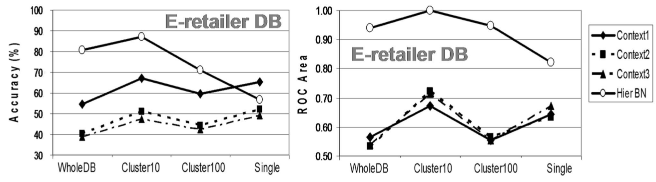
In the case of models built for a single customer, the predictive performance of contextual models is almost always higher than that of the uncontextual model, as Figs. 7 and 8 demonstrate. The percentage of cases in which the difference in performance is negative and statistically significant is around 5 percent for the e-retailer and 20 percent for the food data sets, whereas for the positive and statistically significant case, it is around 75 percent for both data sets, as shown in Fig. 9. Moreover, Figs. 7 and 8 show that the averaged gain in performance obtainable by a personalized contextual versus uncontextual model is higher than 10 percent for the e-retailer data set and is about 4 percent for the food data set. This effect gets blurred when we move from the one-to-one to the *WholeDB* case, because contextual information of individual customers gets lost when we aggregate them into large segments.

5.1.2 The Degree of Contextual Information Matters

Higher values of performance gain can be observed moving to finer degrees of contextual information in Fig. 7b. The gain ranges from 3.5 percent (when K^1 takes two values) to 8 percent (when K^3 takes six values) in the e-retailer data set, in those settings when the contextual models dominate the uncontextual. The performance decreases when the uncontextual model dominates the contextual. However, the loss is moderate, from 1.63 percent to 2.75 percent in absolute terms. The results for the food data set are consistent with the e-retailer data set even if the level of performance is lower: there is 1 percent decay in performance moving to more granular units of analysis, and also the performance improvement moving to finer degrees of contextual information is not relevant. Still the gap in performance between contextual and uncontextual models is statistically significant. This result can be interpreted in terms of the tradeoff between data sparsity and homogeneity. By providing contextual information, customer transactions pertaining to a particular context are reduced making fewer data points fit the model, while the homogeneity of these transactions increases, making the prediction of customer behavior in similar contexts more accurate. The homogeneity induced by the additional contextual information tends to dominate the effect of data sparsity in the one-to-one case. This effect is quantitatively explained in Table 1 (see Section 4.1). The table shows that, as the contextual information becomes finer, the number of transactions selected to learn the predictive model is reduced, but the transactions used in the particular context become more homogeneous, thus providing for better predictive performance of the resulting model.

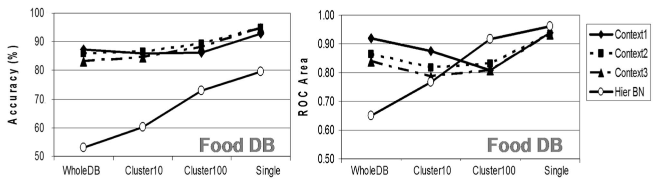
5.2 Can We Infer Context?

The aim is to experimentally demonstrate that it is possible to infer the context in which a transaction took place by comparing the analysis of customers' models in different experimental settings, where the dependent variable is a context attribute, such as a purchase being made for a personal purpose or as a gift. Fig. 10 shows the results for the contextual inference problem obtained by applying models (3a) and (3b) to the e-retailer data set. In Fig. 10a, the performance measure is the predictive accuracy and in Fig. 10b the AUC. The lines labeled as *context1*, *context2*, and *context3* are the results achieved applying model (3a) for each degree of context (K^1, K^2, K^3). The solid line with empty circles (labeled as *Hier BN*) represents performance of model (3b) learned by the BN in Fig. 3. Each line is plotted per degree of customer granularity (x-axis). In both Figs. 10a and 10b, the *Hier BN* model (3b) clearly outperforms each one of the individual contextual models (3a). In particular, the accuracy reaches the maximum value of 90 percent and the AUC of 100 percent when the unit of analysis is *Cluster10*. This demonstrates that inferring the context K^q using the whole hierarchy of contextual information in our experiments provides better results than inferring one context level K^i at a time. Moreover, inferring only a single level does not provide any improvement when the degree of contextual information increases. For instance in Fig. 10a, the inference performance for the coarsest degree of context is higher than for finer levels (*context1* outperforms *context2*



(a) (b)

Fig. 10. E-retailer DB: Performance of contextual inference models for the (a) accuracy and (b) ROC measures.



(a) (b)

Fig. 11. Food DB: Performance of contextual inference models for the (a) accuracy and (b) ROC measures.

and *context3*). 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 point of the *Hier BN* line. With the Friedman test (nonparametric repeated measure ANOVA) [5], we have tested the null hypothesis that the performance of each of the four lines of the graphs is the same. The results are always statistically significant for the accuracy with $p < 0.001$ at least. For the AUC, the difference between the *Hier BN* and each of the remaining models is statistically significant $(p < 0.0001)$, while the difference in performance between each model (3a) is not. The results support the conclusion stated above.

To evaluate the statistical significance of the difference in performance of each point of the *Hier BN* model (e.g., in Fig. 10a, the difference between the 90 percent accuracy in *Cluster10* and the 80 percent in *WholeDB*), we have used the Kruskal-Wallis nonparametric method [4]. The null hypothesis is that the performance obtained by (3b) is the same for each unit of analysis. The differences are always statistically significant with $p < 0.001$ at least.

In Fig. 11, the same inference performance is plotted for the food data set. In this case, model (3a) outperforms model (3b). However, there is no performance improvement when inferring finer degrees of context by model (3a), being the lines very close to each other and the differences in performance not statistically significant. It is quite evident that for the gourmet food data set the results are influenced by the preprocessing performed to grow the gourmet food data set. In fact, the fact that the contextual information was manually introduced according to the principles described in Section 4.1 may constitute a reason why inferring one level of context is easier in the gourmet food data set with respect to the e-retailer data set where contextual data were provided by the customer, not introduced artificially. This is a possible explanation of that the model that infers one context level K^i at a time (3a) performs worse than the model that infers the context K^q using the whole hierarchy of contextual information (3b) in the e-retailer data set, but it is the contrary in the gourmet food case. Fig. 10 can be explained in terms of the heterogeneity versus lack of data tradeoff. As the cluster size decreases, the homogeneity of customers in these clusters increases, *assuming* that clusters are formed using *good* clustering methods. This increase in customer heterogeneity should lead to better predictive performance *assuming* there is enough data to build meaningful predictive models in cases of only few customers and a single customer. If there is enough data to build good predictive models all the way for the single customer case, the performance increases monotonically, as the *Hier BN* line in Fig. 11 clearly attests. If there is not enough data to build good predictive models for small segments and for the single customer case, the performance will drop at the end and there will be a peak in performance, as the *Hier BN* line in Fig. 10 shows. Again, this is the consequence of the customer homogeneity versus not having enough data phenomenon. However, there is also a third case when clustering algorithms produce segments of poor quality and the predictive models on these segments perform poorly, as line *Context1* in Fig. 10b (and some others) attest. Note that the performance improves for *Context1* in Fig. 10b for the single customer case when clustering becomes immaterial.

In summary, we demonstrated in this section that the context can be inferred with good levels of accuracy assuming that "smart" inference methods are used for this purpose. By "smart," we mean that appropriate predictive models combined with the appropriate levels of customer segmentation should be used along with the right performance measures to predict the contextual information, as Figs. 10 and 11 demonstrate. Moreover, given sufficient data, the whole contextual hierarchy can be better inferred than individual contextual variables, as Fig. 10 shows.

5.3 How Do We Exploit Context?

The main goal is to provide the best experimental conditions in which contextual information can be utilized for predicting users' behavior. This is important because the contextual knowledge should improve such predictions, which makes it necessary to identify the best possible conditions. This issue can be analyzed in two ways: 1) once you *know* the context, how can it be exploited in the best possible manner and 2) if you *do not know* context, how can it be inferred and then used to improve model performance. We explore these two questions below.

How to utilize existing contextual information. To this aim, two context-dependent customer models (shown on the right in Fig. 3) have been compared to the uncontextual model in different experimental settings. The first model is the standard Naïve Bayes model (4a), where the context is captured with an independent attribute K^q. The second contextual model is a BN (4b), where the contextual information is modeled as a variable K^q in the middle layer of the network, as shown in Fig. 3. The performance of each of the contextual models (4a) and (4b) is compared with the performance of the basic uncontextual model (1) across various experimental setting using the $Diff$ measure defined by (5). Figs. 12a and 12b show the differences in performance between contextual and uncontextual models for the e-retailer and gourmet food data sets, respectively. The length of the black column is the average difference in

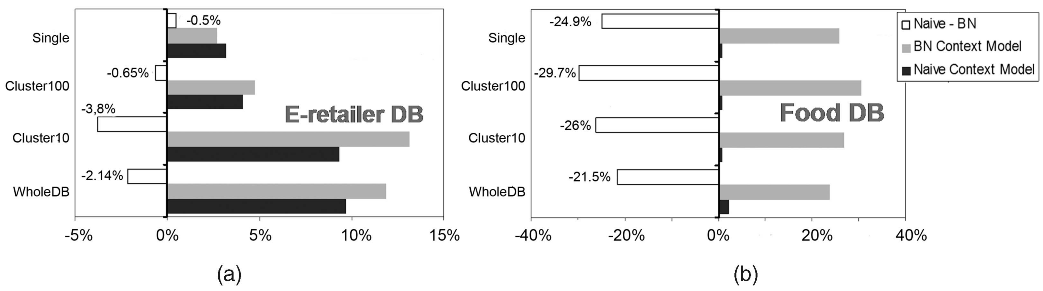


Fig. 12. Difference in performance of contextual and uncontextual models per customer granularity for the (a) e-retailer and (b) food data sets.

performance $Diff$ between the *naïve (Naïve Bayes)* contextual model (4a) and the uncontextual model (1) taken across various experimental settings, such as types of dependent variables used in these models and types of performance measures. The gray column is the average difference $Diff$ between the Bayesian contextual model (4b) and the uncontextual model (1). The white column is the average difference $Diff$ between the *naïve* contextual model (4a) and the *Bayesian* contextual model (4b). As Fig. 12a shows, the *Bayesian* model (4b) is always better than the *naïve* model (4a), except when the unit of analysis is a single customer. Each column is plotted for each level of market granularity (vertical axis) ranging from the whole customer base (*WholeDB*) to the individual models of customers (*Single*). For example, in the e-retailer data set, when the unit of analysis is *Cluster100* the Bayesian contextual model (gray column) is almost 5 percent better than the uncontextual. As Fig. 12a demonstrates, the *naïve* contextual (4a) and the *Bayesian* contextual (4b) models always outperform the uncontextual model (1) for the e-retailer data set. Moreover, these performance differences are statistically significant. The null hypothesis is that the performance of each model is the same [(4a) versus (1) and (4b) versus (1)]. The differences are statistically significant with $p < 0.001$ for both the naïve and Bayesian models and for all the levels of market granularity. For the gourmet food data set in Fig. 12b, the performance differences between the Bayesian (4b) and the uncontextual (1) models (the gray bars in Fig. 12b) are also statistically significant (at least $p < 0.001$) but not for the naïve case ($p > 0.05$). The variation in the relative performance of the two contextual models (shown by the white column in Fig. 10a) ranges from -3.8 percent to $+0.5$ percent and can be explained as follows: The *Bayesian* model is "data hungry," meaning that in order to leverage the contextual heterogeneity and improve the predictive performance, a certain amount of data is needed per unit of analysis. Therefore, the *Single* customer level is suboptimal (not enough data), as well as the *WholeDB* level (the context gets lost). Thus, the biggest performance difference is in between these two extremes. For e-retailer data, the best market granularity level is *Cluster10*. For the gourmet food data set, the performance difference between (4a) and (4b) is even bigger (the white column is far to the left of the y-axis) and statistically even more significant. Note that the $Diff$ measure reaches the peak at *Cluster100* (-29.7 percent), rather than *Cluster10*, as in the e-retailer case.

The results pertaining to the first question (Does context matter?) can be summarized as follows:

There is a significant improvement in the predictive performance when the contextual information is modeled as a hidden variable (in a BN) versus using it as an independent variable (in an NB model). This conclusion is in line with the result reported in [30] that a BN can outperform the Naïve Bayes approach given the same set of attributes.

Limited data cannot leverage the heterogeneity effect. There is a specific degree of customer granularity, which improves the predictive performance depending on the inner characteristics of the data set. Given that the Bayesian contextual approach outperforms the naïve contextual model, one needs to determine the best degree of customer granularity. In both data sets, there is statistical evidence that the solution is in the middle of the customer granularity scale. It means that even if the firm is able to work on each single customer, not necessarily the best performance results will be obtained because of the variance bias problem and the complexity of the BN that needs more data compared to the naïve approach. In each unit of analysis, there is contextual heterogeneity, but in some cases, there is not enough data to let the model learn from the limited available data. In the e-retailer data set, there is an average of 40 transactions per customer that rises to 200 for *Cluster100*. In this case, the best market granularity level is *Cluster10* since for *Cluster100* there is not enough data to capture the contextual heterogeneity, whereas heterogeneity is too high for the whole data set. In the food data set, the results are similar; but the best unit of analysis is *Cluster100* rather than *Cluster10* because of the size of the data set: the average number of transactions for *Cluster100* is about 700, which is enough to capture the context effect while keeping heterogeneity "under control."

How to utilize inferred contextual information. As was argued in Section 5.2, if done properly, the contextual information can be successfully inferred. In this section, we show that this inference can provide for better predictions of customer behavior. To do this, we use the latent BN model (4c) (that is also shown in Fig. 3 as the rightmost diagram) that represents the unknown contextual information with the latent variable K^q and compare its performance with the performances of the uncontextual model (1) and the contextual BN model (4b). In other words, we do this comparison for a BN model across the three levels of context: no context [model (1)], full context [model (4b)], and latent context [model (4c)]. The uncontextual model (1) is defined with the NB model constituting a simple case of a

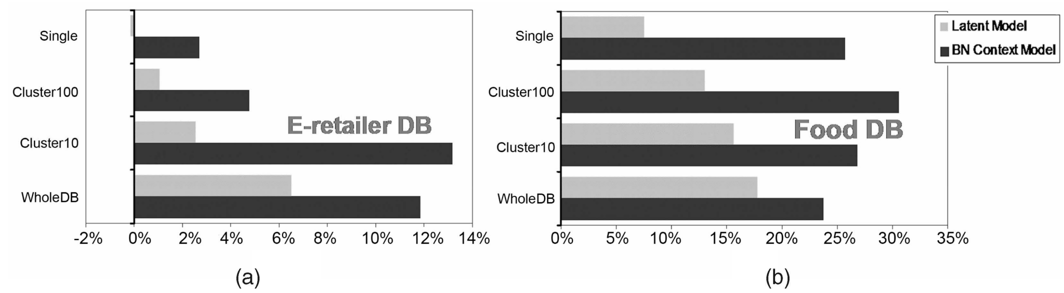


Fig. 13. Difference in performance of latent and Bayesian contextual models per customer granularity for (a) e-retailer and (b) food data sets.

BN, while models (4b) and (4c) are defined with the same BN in Fig. 3; but (4b) uses available contextual information, whereas model (4c) infers it instead. The performance of models (4b) and (4c) is compared with that of the basic uncontextual model (1) across various experimental settings using the $Diff$ measure defined by (5).

Figs. 13a and 13b show the results for the e-retailer and gourmet food data sets, respectively. The length of the black column is the average difference in performance $Diff$ between the *Bayesian* contextual model (4b) and the uncontextual model (1) taken for various experimental settings, such as types of dependent variables used in these models and types of performance measures. The gray column is the average difference $Diff$ between the latent (4c) and the uncontextual (1) models. Each column is plotted for each level of market granularity (vertical axis) ranging from the whole customer base to the individual level. For example, in the e-retailer data set, when the unit of analysis is *Cluster100* the latent model (gray column) is almost 1 percent better than the uncontextual. As Fig. 13a demonstrates, the *Bayesian* contextual (4b) and the *latent* (4c) models always outperform the uncontextual model (1) for both data sets. The differences are statistically significant except in the single user unit of analysis (e-retailer data set) where the latent model has a small negative performance. To show this, we formulated the null hypothesis: the performance of each model is the same [(4b) versus (1) and (4c) versus (1)]. The performance differences are statistically significant (with $p < 0.001$) for all comparisons and levels of market granularity, except for the previous case. As expected, the latent models in all the experimental conditions are worse than the Bayesian contextual model where the context is explicitly known and those differences in performance are statistically significant (at least $p < 0.001$). From all this, we conclude that the latent model (4c), inferring the context, outperforms the uncontextual model (1) but is dominated by the similar contextual model (4b). This means that even without explicit knowledge of context, we can still infer the context and use it for better predicting customer behavior, although the performance is lower (from 3 percent to 10 percent in the e-retailer and from 7 percent to 22 percent in the food data set) than when explicit knowledge of context is available.

The results reported in this section have significant implications for marketers and data miners. They demonstrate that nothing beats explicit contextual knowledge; but if acquiring this knowledge is expensive, the second best alternative is to try to infer and use it appropriately.

6 CONCLUSIONS

In this paper, we have addressed three questions: 1) does context matter when building customers' behavioral models, 2) to which extent is it possible to infer the contextual information from the data, and 3) how do we use the contextual information for modeling user behavior.

For the first research question, we conclude that, first, context matters in the case of modeling the behavior of *individual* customers, i.e., knowing the context in which a customer makes the purchase improves the ability to predict that customer's behavior. Second, the degree of contextual information also matters: the more we know about the context of a transaction, the better we can predict the customer's behavior. Third, the effect of contextual information gets diluted during the process of aggregating customers' data. Context does matter for individual customers but does not significantly matter when predicting the behavior of the whole customer base. Also, the contextual effects are stronger for the finest levels of customer segmentation. All these results can be interpreted in terms of the tradeoff between data sparsity and customer homogeneity. By providing contextual information, customer transactions pertaining to this particular context are reduced, making fewer data points to fit the context-specific model, while homogeneity of these transactions increases, making it easier to predict more accurately customer behavior in similar contexts. As shown in Section 5, the number of transactions available to the model to learn the predictive function can drop from more than 5,000 to less than 10 (in the e-retailed data set and finest contextual degree), but the gain in the predictive performance can increase from 6 percent to 10 percent, witnessing the increase in the homogeneity of the customer's behavior.

For the second question (can we infer contextual information?), we conclude that it is possible to do this for specific levels of customer segmentation using the best-of-breed predictive methods. These "specific segmentation levels" depend on several parameters discussed in this paper and are dictated by the optimal amounts of pooled contextual customer data necessary to make these inferences. As our results show, the BN approach captures inner dependencies between the attributes of the model and the underlying context, leading to good predictions and contextual inferences. For example, a BN (shown in Fig. 3) inferred the purchasing context with 90 percent to 100 percent performance rate in the e-retailer data set, as demonstrated in Fig. 10, thus making it unnecessary to explicitly collect contextual information in those cases. This inference needs to be done for the right level of customer segmentation since both inference and predictive performances peak at a certain level of customer segmentation. The contextual inference process is reliable only if the analyst can properly select a good model (such as a certain type of a BN for our case) and identify a proper segmentation level.

For the third research question (how do we utilize the contextual information?), we conclude that, first, knowing context is not enough and that exploiting it in the best possible way is also very important. For example, incorporating the context into a BN produces better prediction results than simply using it as a part of NB. Dropping the independence condition between the contextual information and some of the independent variables of the model can be a useful solution for improving predictions. Second, there is a significant predictive performance improvement when the contextual information is modeled as a latent variable in a "smart" way. This demonstrates that if the contextual information cannot be acquired from external sources, it can still be used as a latent variable and contribute to the overall improvement of predicting customers' behavior. This result has important practical implications for marketers and data miners dealing with e-commerce applications since acquiring contextual information can be costly. Third, there is a specific customer segmentation level that maximizes predictive performance, as shown in Figs. 12 and 13, which depends on the inner characteristics of the data set.

The results cannot be generalized to every data set and to all industry sectors, but they can be generalized to a relevant number of e-commerce applications, especially to the "high-frequency" applications where customers transact frequently and in the contexts defined by certain hierarchical taxonomies. The customer behavior is very likely to change in different contexts and we tried to understand these changes better in this study. Moreover, several e-commerce applications are already structured to capture the intent of a customer's purchase by explicitly asking the user whether he/she is going to make a gift. In these cases, this research results should turn out to be interesting for managers. The caveat for firms and managers is that caring about context is not enough: they have to carefully consider in which conditions the contextual information gathered is going to be used and for which marketing decisions.

ACKNOWLEDGMENTS

This paper is based on the preliminary research published in [14] and [24].

REFERENCES

[1] G. Adomavicius, A. Tuzhilin, R. Sankaranarayanan, and S. Sen, "Incorporating Contextual Information in Recommender Systems Using a Multidimensional Approach," *ACM Trans. Information Systems,* vol. 23, no. 1, pp. 103-145, 2005.

[2] G. Adomavicius and A. Tuzhilin, "Multidimensional Recommender Systems: A Data Warehousing Approach," *Proc. Second Int'l Workshop Electronic Commerce (WELCOM '01),* vol. 2232, pp. 180-192, 2001.

[3] G. Akrivas, M. Wallace, G. Andreou, G. Stamou, and S. Kollias, "Context-Sensitive Semantic Query Expansion," *Proc. IEEE Int'l Conf. Artificial Intelligence Systems (ICAIS '02),* pp. 104-109, 2002.

[4] J.W. Barnes, *Statistical Analysis for Engineers and Scientists.* McGraw Hill, 1994.

[5] A. Berry and G. Linoff, *Data Mining Techniques: For Marketing, Sales, and Customer Support.* Wiley Computer Publishing, 1997.

[6] J.R. Bettman, M.F. Luce, and J.W. Payne, "Consumer Decision Making: A Constructive Perspective," *Consumer Behavior and Decision Making,* M. Tedeschi, ed., pp. 1-42, 1991.

[7] C. Bothorel and K. Chevalier, "How to Use Enriched Browsing Context to Personalize Web Site Access," *Proc. Sixth Int'l Conf. Modeling and Using Context (CONTEXT '03),* pp. 419-426, 2003.

[8] P. Brézillon, "Context in Artificial Intelligence: I. A Survey of the Literature," *Computer and Artificial Intelligence,* vol. 18, no. 4, pp. 321-340, 1999.

[9] P. Brézillon, "Task-Realization Models in Contextual Graphs," *Proc. Fifth Int'l and Interdisciplinary Conf. Modeling and Using Context (CONTEXT '05),* Lectures Notes in Artificial Intelligence, vol. 3554, pp. 55-68, 2005.

[10] P.J. Brown, J.D. Bovey, and X. Chen, "Context-Aware Applications: From the Laboratory to the Marketplace," *IEEE Personal Comm.,* vol. 4, no. 5, pp. 58-64, 1997.

[11] G. Chen and D. Kotz, "A Survey of Context-Aware Mobile Computing Research," Technical Report TR-2000-381, Dept. Computer Science, Dartmouth College, 2000.

[12] A.K. Dey, G.D. Abowd, and D. Salber, "A Conceptual Framework and a Toolkit for Supporting the Rapid Prototyping of Context-Aware Applications," *Human Computer Interaction,* vol. 16, no. 2, pp. 97-166, 2001.

[13] D. Franklin and J. Flaschbart, "All Gadget and No Representation Makes Jack a Dull Environment," *Proc. AAAI Spring Symp. Intelligent Environments,* 1998.

[14] M. Gorgoglione, C. Palmisano, and A. Tuzhilin, "Personalization in Context: Does Context Matter When Building Personalized Customer Models," *Proc. Sixth IEEE Int'l Conf. Data Mining (ICDM),* 2006.

[15] D. Heckmann et al., "GUMO—The General User Model Ontology," *Proc. 10th Int'l Conf. User Modeling (UM),* 2005.

[16] T. Jiang and A. Tuzhilin, "Segmenting Customers from Population to Individuals: Does 1 to 1 Keep Your Customers Forever," *IEEE Trans. Knowledge and Data Eng.,* vol. 18, no. 10, pp. 1297-1311, Oct. 2006.

[17] P. Kotler, *Marketing Management.* Prentice Hall, 2003.

[18] S. Lawrence, "Context in Web Search," *IEEE Data Eng. Bull.,* vol. 23, no. 3, Sept. 2000.

[19] D.B. Leake and R. Scherle, "Towards Context-Based Search Engine Selection," *Proc. Sixth Int'l Conf. Intelligent User Interfaces (IUI),* 2001.

[20] G.L. Lilien, P. Kotler, and S.K. Moorthy, *Marketing Models,* pp. 22-23. Prentice Hall, 1992.

[21] X. Luo, "The Performance Implications of Contextual Marketing for Business-to-Consumer Electronic Commerce: An Empirical Investigation," *J. Database Marketing,* vol. 10, no. 3, pp. 231-239, 2003.

[22] J.G. Lussier and R.W. Olshavsky, "Task Complexity and Contingent Processing in Brand Choice," *J. Consumer Research,* vol. 6, no. 2, pp. 154-165, 1979.

[23] Z. Maamar, D. Benslimane, and N.C. Narendra, "What Can Context Do for Web Services?" *Comm. ACM,* vol. 49, no. 12, 2006.

[24] C. Palmisano, A. Tuzhilin, and M. Gorgoglione, "User Profiling with Hierarchical Context: An E-Retailer Case Study," *Proc. Sixth Int'l and Interdisciplinary Conf. Modeling and Using Context (CONTEXT '07),* Aug. 2007.

[25] A. Pappas, S. Hailes, and R. Giaffreda, "A Design Model for Context-Aware Services Based on Primitive Contexts," *Proc. First Int'l Workshop Advanced Context Modelling, Reasoning and Management,* 2004.

[26] C.K. Prahalad, "Beyond CRM: C.K. Prahalad Predicts Customer Context Is the Next Big Thing,," *AMA MwWorld,* 2004.

[27] J.F. Rayport and J.J. Sviokla, "Managing in the Marketspace," *Harvard Business Rev.,* vol. 68, no. 6, pp. 141-150, 1994.

[28] T. Rodden, K. Cheverst, K. Davies, and A. Dix, "Exploiting Context in HCI Design for Mobile Systems," *Proc. First Workshop Human Computer Interaction with Mobile Devices (Mobile),* 1998.

[29] N. Ryan, J. Pascoe, and D. Morse, "Enhanced Reality Fieldwork: The Context-aware Archaeological Assistant," *Computer Applications in Archaeology.* Oxford, 1997.

[30] P. Sarkar, G. Nagy, and S. Veeramachaneni, "Modeling Context as Statistical Dependence," *Proc. Fifth Int'l Conf. Modeling and Using Context (CONTEXT),* 2005.

[31] B. Schilit and M. Theimer, "Disseminating Active Map Information to Mobile Hosts," *IEEE Network,* vol. 8, no. 5, pp. 22-32, 1994.

[32] H.R. Schmidtke, "Granularity as a Parameter of Context," *Proc. Fifth Int'l Conf. Modeling and Using Context (CONTEXT),* 2005.

[33] A. Sieg, M. Bamshad, and R. Burke, "Representing Context in Web Search with Ontological User Profiles," *Proc. Sixth Int'l and Interdisciplinary Conf. Modeling and Using Context (CONTEXT '07),* Aug. 2007.

[34] D. Vallet, M. Fernandez, P. Castells, P. Mylonas, and Y. Avrithis, "A Contextual Personalization Approach Based on Ontological Knowledge," *Proc. 17th European Conf. Artificial Intelligence (ECAI)—Contexts and Ontologies: Theory, Practice and Applications Workshop (C&O),* 2006.

[35] A. Ward, A. Jones, and A. Hopper, "A New Location Technique for the Active Office," *IEEE Personal Comm.,* vol. 4, no. 5, pp. 42-47, 1997.

[36] H. Zhu and S. Madnick, *Structured Contexts with Lightweight Ontology,* MIT Sloan Research Paper 4620-06, 2006.

[37] C. Ziegler, M. McNee, J. Konstan, and G. Larson, "Improving Recommendation Lists through Topic Diversification," *Proc. 14th Int'l World Wide Web Conf. (WWW '05),* May 2005.

Cosimo Palmisano received the Master's degree in electronic engineering and the PhD degree in management engineering from the Polytechnic of Bari, Italy. He is senior business data analyst consultant at the Customer Intelligence Department of FIAT Group Automobiles. He has been a visiting scholar in the Information and Operations Management System Department, Stern School of Business, New York University. His current research interests include data mining, targeted sales actions, customer relationship management, recommender systems, and utilizing contextual information in segmentation problems. He has published numerous papers in various information systems and management international conferences. He has been involved in several industrial research projects related to the above-mentioned research fields. He is a member of the IEEE and the IEEE Computer Society.

Alexander Tuzhilin received the PhD degree in computer science from the Courant Institute of Mathematical Sciences, New York University (NYU). He is a professor of information systems and the NEC faculty fellow in the Stern School of Business, NYU. His current research interests include data mining, personalization, recommender systems, and CRM. He has published widely in the leading CS and IS journals and conference proceedings on these and other research topics. He served on the organizing and program committees of numerous CS and IS conferences, including as a program cochair of the Third IEEE International Conference on Data Mining. He has also served on the editorial boards of the *IEEE Transactions on Knowledge and Data Engineering*, the *Data Mining and Knowledge Discovery Journal*, the *INFORMS Journal on Computing* (as an area editor), the *Electronic Commerce Research Journal*, and the *Journal of the Association of Information Systems*. He is a member of the IEEE and the IEEE Computer Society.

Michele Gorgoglione received the PhD degree in business engineering from the University of "Roma-Tor Vergata," Rome. He is an associate professor at the Polytechnic of Bari, Italy, where he teaches marketing, financial accounting and management, and e-business models in the College of Engineering. His current main research interest is on data mining models for customer profiling, particularly in the use of context to improve customer modeling and business performance. This research is done in collaboration with the New York University. He has also done research on creativity support systems, on knowledge management, and on the use of neural networks and fuzzy logic systems to support decision-making processes in companies. He is a member of the IEEE and the IEEE Computer Society.