D.Sailaja, K.Nasaramma, M.Sumender Roy, Venkateswarlu Bondu / International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 www.ijera.com
Vol. 2, Issue 2,Mar-Apr 2012, pp.1134-1138

Predictive Modeling of Customers in Personalization Applications with Context

D.Sailaja, K.Nasaramma, M.Sumender Roy, Venkateswarlu Bondu

Nasaramma.K is currently pursuing her M.Tech in Godavari Institute of Engineering and Technology. Rajahmundry, India
Sailaja.D is currently pursuing her M.Tech in Avanti Group of Colleges, Visakhapatnam, India.
Sumender Roy is currently working as Professor in Godavari Institute of Engineering and Technology. Rajahmundry, India.
Venkateswarlu Bondu pursuing Ph.D in Computer Science in Andhra University.

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 project, we show 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. 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 make context valuable for improving predictive performance of customers' behavior.

Index Terms— Personalization, Context, Data Mining, User Modeling, Predictive Modeling.

1. INTRODUCTION:

1.1 Introduction to Project:

Contextual information indeed makes a significant difference in building better customer models in marketing and e-commerce applications. In this project, 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. Therefore the research question that we just described can be summarized as follows:

Does context matter for building better models to predicting customer behavior?

In this paper, we answer the question empirically by conducting an empirical study on data set across a wide range of experimental conditions. To answer the question, we built two alternative customer models, one including contextual information and the other one not, and compared their predictive performances. This study makes the following contributions to studying context in personalization applications. First, we demonstrate that context indeed matters when predicting customer behavior for whole or small homogenous groups of customers and gets diluted during the process of aggregating customers' data. Finally, the context is taking externally, and then used for predicting customer's behavior. We show that the resulting model significantly outperforms the basic uncontextual model.

1.2 Introduction to modules:

1.2.1 Collecting dataset externally.
1.2.2 Clustering the data.
1.2.3 Apply predictive modeling on datasets and clusters.

1.2.1 Collecting dataset externally:

The experiment has been conducted on e-retailer dataset which is created by us. For each customer, the following demographic data were added: age, previous studies, marital status, composition of the family, place of living, hobbies, and whether the customer owned a car. The transactional data include: item purchased, price, day, time, session duration, number of clicks per connection, and the time elapsed for the web page.

1.2.2 Clustering the data:

Apply K-mean clustering algorithm on dataset and divide into clusters: cluster1, cluster2 and cluster3.

1.2.3 Apply predictive modeling on dataset and clusters:

Apply Naive-Bayes classification algorithm to predict customer behavior on whole dataset and each clusters are cluster1, cluster2 and cluster3.

2. PROBLEM FORMULATION:

In our project, first explain what we mean by “context,” then how we model customer behavior, and finally, the methodology for comparing contextual and uncontextual models.

2.1 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 something.” In the data mining community, context is defined 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. Other factors have been added to the previous definition. For instance, includes the date, the season, and the temperature. Add the physical and conceptual statuses of interest for a user and 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, while others emphasize how context relates to the application? 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. 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 K that can have a complicated structure reflecting the complex nature of this information. Each contextual attribute K in K is defined by a set of q attributes: K = (K1; . . .;Kq) contextual attribute K specifying the intent of a purchasing transaction in an e-retailer application, considered. K is personal context: personal purchase made for the work-related or other purposes. Similarly, the Gift value for K can be split into a gift for a partner or a friend and a gift for parents or others. Thus, K= {PersonalWork; PersonalOther; GiftPartner/Friend; Gift Parent/ Other}. Finally, attribute K to be taken.

2.2 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; . . .;A_m}, and a set of r transactions Trans(C_i)={ TR_{i1}; TR_{i2}; . . . ; 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; . . . ; T_p}. In addition, we also have contextual information K associated with each transaction TR_{ir}, in Fig.1 represents 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}. In general, however, we support multiple contextual attributes. Finally, the customer base C can be partitioned into several segments by computing summary statistic S for customer C_i over the transactions Trans(C_i)={TR_{i1}; TR_{i2}; . . . ; TR_{ik}} made by that customer using statistical aggregation and moment function, such as mean. For instance, for the transactions made by the customers in the previous example, the statistic can be S = {Average price}. This means, among other things, that each customer C_i has a unique summary statistic S 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 space of statistics S, customers can be clustered into groups (segments) in that space using the K-Means clustering technique. 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 $Trans(C_i)=\{ TR_{i1}; TR_{i2}; \ldots ; TR_{ik}\}$ for customers i in p_α, we want to build a single predictive model M_α on this segment of customers p_α:

$$Y = f(X_1;X_2; \ldots;X_p); \ldots\ldots(1);$$

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 real, i.e., $f(p_\alpha) \in R$. 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. The predictive models do not assume any contextual information since the contextual variable K is not a part of the model. Therefore, we call the model of this type uncontextual. We define contextual counterparts of predictive model (1), where the model takes the following form:

	Demographic Attributes A			**Transactional Attributes T**			**Context K**		
	A_1	…	A_m	T_1	…	T_p	K_1	…	K_q
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}	$T_{jr.1}$	…	$T_{jr.p}$	…	…	…

$$Y = f_{Kq=a}(X_1;X_2; \ldots;X_p); \ldots\ldots(2)$$

Where the model 2 constitute the way of creating a contextual model. Model 2 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 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 .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 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 research question (Does context matter?), a comparison between performance results of predictive models is performed for the uncontextual (1) and the contextual (2) models across a wide range of experimental conditions can be specified by replacing the dependent variable Y in (1) with the context variable K as

$$K_q = f(X_1;X_2; \ldots;X_p); \ldots\ldots\ldots\ldots(3).$$

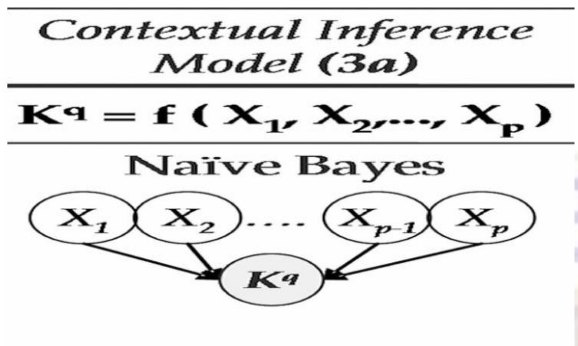
In model (3), the dependent variable is the contextual information. One contextual attribute is inferred at a time. For model (3), f is a predictive function learned via naïve bayes model.

2.3 Figures:

Demographic, transactional, and contextual information about the customers and their transactions.: ***Fig. 1***

Graphical representation of the predictive model:

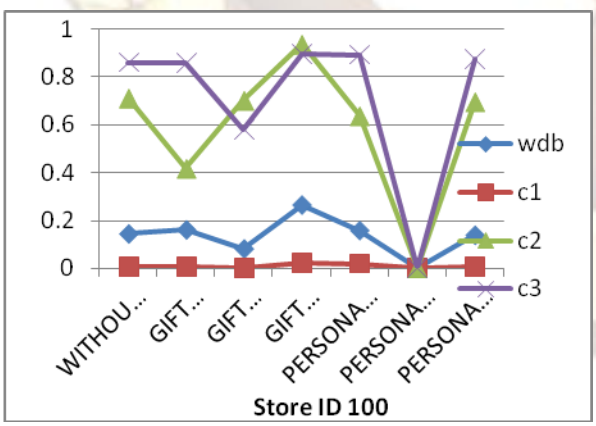
Fig. 2.

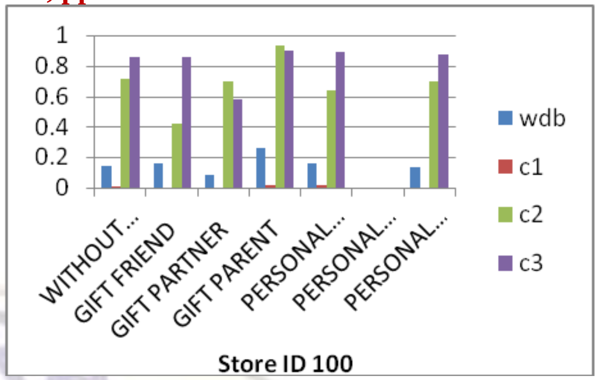


3. Results:

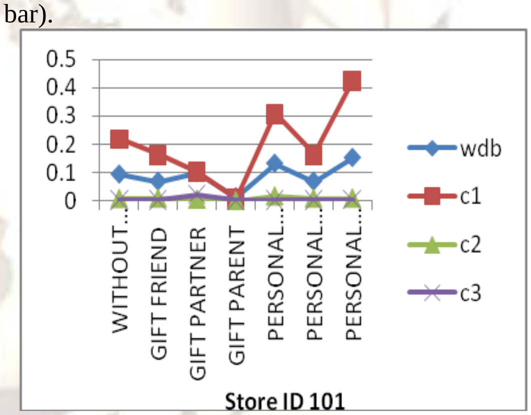
3.1 Does Context Matter?

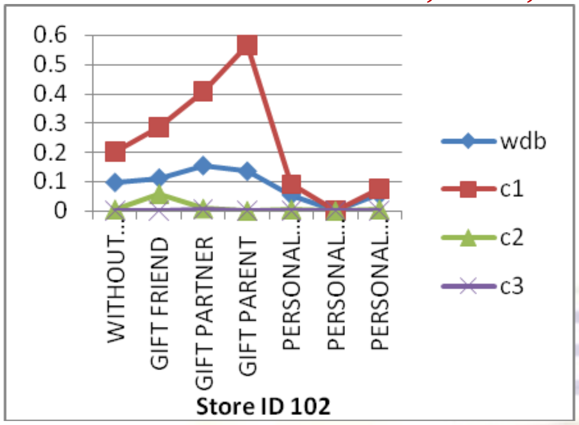
To give a "flavor" of the results, given the experiment settings (contextual information, data set, one classifier, two dependent variables)., the below figures presents three graphs generated by plotting the values of the e-retailer data set for contextual information and uncontextual. The graphs are presented in the order of progressively more refined contextual information. Moreover, most of the contextual models show a better predictive performance compared to the uncontextual, except for some cases where the difference is very small. Although for these charts the curves are not always monotonic, the predictive performance of the contextual models is usually higher than that of the uncontextual model.





Analyzing and comparing the performance curves, the first evident result shows the un-contextual model is always below the other two contextual models. It means that, independently from the level of analysis, and from the experimental settings, gathering contextual information gives better results in terms of customers' behavior predictions. Another interesting point is the monotonic shape of each performance curve. The results shows better results in context compared to uncontextual in three different stores (store100, store101, and store102). In three stores shows context is better compared to uncontext in two types of charts (curve, bar).





4. CONCLUSION & FUTURE WORK:

4.1 Conclusion:

This work has been conducted a comparative study of contextual and un-contextual models across multiple dimension of analysis such as the context variable, classifier algorithm and different types of metrics. This analysis has been performed with the aim of demonstrating the relevance of the context in building customer profiles.

Our results show that, in different experimental settings, the contextual model, in general, outperforms the un-contextual one; this overcome is more relevant. This underlines the fact that the context matters in building customers' purchasing profiles and it should be of great effectiveness if applied for building personalized recommendation in the e-commerce environment.

Further research will be held making more complex the experimental settings in order to give more empirical evidence of the results. And will prove two more questions for better predicting

- ✓ Is it necessary to acquire contextual information or is it possible to infer it from the data?
- ✓ How do we exploit the inferred contextual information for modeling customer behavior?

The experiments will be conducted predicting different transactional variable *Y*, using more classifier algorithms *f*, considering more levels of market granularity through finer segments, and generating more contextual models using the further sublevels of the "personal" and "gift" shopping mode variables.

5. BIBOLOGRAPHY:

1. Using Context to Improve Predictive Modeling of Customers in Personalization Applications. – Cosimo Palmisano, Alexander Tuzhilin, Michele Gorgoglione. IEEE Transaction on Knowledge and Data Engineering.

2. Data Mining – Concepts and Techniques.- Han & Kamber.

3. Increasing Customer Value by Integrating DataMining and Campaign Management Software, Direct Marketing Magazine, Feb, 1999.

4. Data Mining: Introductory and Advanced Topics – Margaret Dunham. Prentice Hall, 2003.

5. Mining the e-commerce data to analyze the target customer behavior. – Yuantao Jiang, Siqin Yu.

6. Java Complete Reference. Herbert Schildt.

7. Unified Modeling Language. Grady Booch.

8. http://www.wikipedia.org