Personalization and Location-based Technologies for E-Commerce Applications

K. V. Ravi Kanth, and Siva Ravada
Spatial Technologies, NEDC,
Oracle Corporation, Nashua NH 03062.
{Ravi.Kothuri, Siva.Ravada}@oracle.com

Abstract

Tailoring web-pages to different user characteristics such as location, preferences and previous history (page-hits, products bought) have been shown to be effective tools for personalizing web-content. In this paper, we briefly summarize the techniques in these state-of-the-art personalization technologies. We first describe personalization using user preferences or history and then describe personalization based on user's current location. Whereas the former is applicable for deployment in web-sites, the latter is useful in providing location-based content to mobile users and wireless applications.

1 Introduction

In the traditional business industry, database marketing enables the identification of prospective customers who are likely to respond to mass product mailings. Instead of randomly mailing product listings to prospective customers, database marketing uses parameters such as demographics, recent purchase history, and customer profiles. In the case of e-businesses, similar techniques can be applied using personalization techniques. Purchase history is kept track by the business web-sites, products that could be interesting to the user are kept track using the page-hit history, along with user-profile characteristics such as location, and income. Several e-businesses employ such personalization techniques for successful marketing: Amazon.com points a customer to related items whenever the customer searches for a specific item and accesses the corresponding web-page. Likewise demographics can be used to successfully infer user's characteristics. For example, Carrier Corp. of Connecticut uses customer's zip-code to market specific products to customers. The zip-code and other address data are used to infer whether the customer comes from an affluent neighborhood or not and product recommendations are then inferred.

In some cases, the address of the customer may not be of importance. Instead, the current location of the user may be used for personalization. This is especially true in the case of mobile users. The wireless service provider automatically identifies the location of the user and could provide dynamically rendered web content showing places and features of interest at that specific location. For example, ads on movie theaters, restaurants and other common places of interest may be displayed whenever the customer uses the mobile device to retrieve stock quotes or other information. In what follows, we first summarize the techniques behind personalization technology and then describe how location can be used both for mobile and non-mobile e-customers.

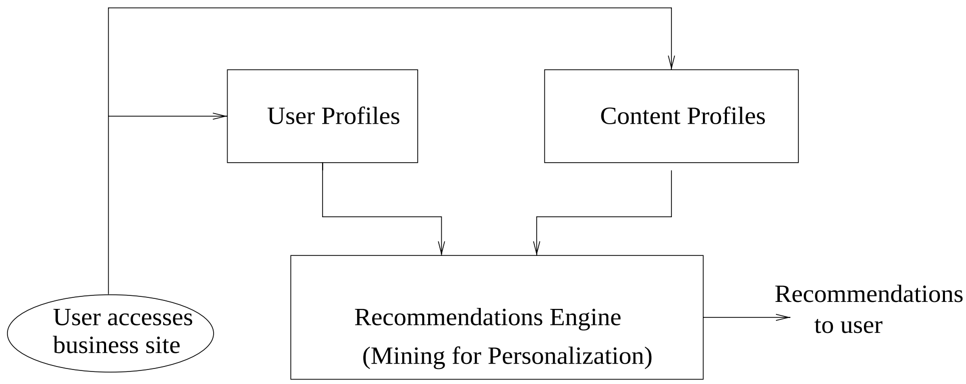

Figure 1: Personalization Components for E-commerce.

2 Personalization for E-commerce

In this section, we present a general overview of the e-business architecture to enable personalized marketing. Detailed explanations and models are presented in [4]. Most E-businesses maintain three components: (1) User profiles, (2) Content Categories, and (3) Business rules as depicted in Figure 1. We describe each of these in more detail.

User Profiles User profiles capture the characteristics/interests of the customer. These typically include demographics such as city, address of residence, job profile, specific interests etc. This information could be leveraged to categorize a new customer and lead him to a tailored-website for that user-profile. In addition, user profiles could be implicitly updated based on the user behavior e.g., pages/products accessed by the user etc. In some cases, the users may not register with the business and hence there may not be have any previous history. Such users are referred to as visitors (whereas others are referred to as customers). The recommendations for visitors are made using the information from the current session and any demographic data, if available.

Content Profile Content profiles keep all data about the content: this includes the information associated with the content (e.g., the composer, singer for a song item) as well as the business value for the content such as the number of items in stock, number of times the product is considered (website accessed) and the number of items for this product sold. These profiles enable analysis by the business and allow them to identify whether a product needs to be discounted, or if the market has caught up for the item (or if the item is under-stock) if it is be removed from the discounted list.

Business Recommendations/Rules Engine Using the user profiles and content profiles, the businesses apply data mining techniques [10, 11] to identify appropriate business rules. These rules could involve a simple classification of the users using their profiles and the website click-streams, association between content profiles and user behavior, or association between different products (e.g., customer buying "Concrete Mathematics" is highly likely to also buy "The Art of Computer Programming"). In what follows, we describe these data mining techniques in more detail.

3 Data Mining for Personalization

Classical data mining techniques include classification of users, finding associations between different product items or customer behavior, and clustering of users. Implicit concept generalization techniques such as updating the content profiles for a brand name, when a customer buys a product from that brand name also need to be applied. Details on this can be found in [] and will not be dealt with here. Next, we describe the main data mining techniques for use in personalization.

3.1 Clustering

Given a set of users (or profiles), and their buying patterns, clustering combines the users into clusters which can later be used for prediction and personalization for prospective customers. Each product represents a separate dimension and the preferences of each user (for all the products) constitute a point-vector in a multi-dimensional product space. Identifying groups of users that select similar products then transforms to finding closely-knit clusters in the multi-dimensional product space. Several algorithms such as k-mean, k-medoids, CLARANS, and BIRCH have been proposed for such clustering. If the number of products is high, the corresponding high-dimensional clustering techniques could be ineffective. For such cases, dimensionality-reduction techniques [13, 9, 3] could be employed.

3.2 Classification

Classification categorizes the customers based on their user profiles and their product-access behavior. A set of pre-classified customers are used to train the classification network. Classification is performed using one of the following popular techniques: decision trees, k-nearest neighbor, or neural networks. Most commercial data mining engine vendors such as Oracle and IBM provide on or more of these techniques.

3.3 Association Rules

Association rules identify the interest (purchase of access) in more than one item in the same transaction. For example, students buying one book are also highly likely to buy another since both the books are more complimentary in most graduate course studies. An association rule is a condition of the form X implies Y where X and Y are two product items. The *support* for the rule is the fraction of transactions that contain both X and Y and the *confidence* for the rule is the fraction of transactions containing X, which also contain Y. The business sets the support and confidence and the underlying mining engine identifies all interesting associations that satisfy the specified support and confidence measures. Such identification is performed using several well-known algorithms such as *a priori* [18, 17] etc. The identified associations can be then used to market item Y whenever a customer buys item X.

4 Location in E-commerce

Location plays a vital role in both personalization (as described above) as well as an enabling component. As an enabling component, location is used to identify interesting stores in his location using a mapquest-like locator service. For example, a customer could identify the nearest store sites within a specified distance from his current location. More details on this aspect could be found in recent spatial database research [1]. Next, we examine the role of location in personalization and wireless e-commerce.

For most businesses, customer location is a frequently used attribute in personalization of the content. For example, customers from the Southern states may have special preferences such as country music CDs. To detect such hidden relationships between location and customer behavior, location could be explicitly requested from the customer at the time of registration (or implicitly inferred from the customer's behavior). Given a training sample of customers, location categorized to zip-code, city, county, and state levels can be used to classify customers [7, 6, 5]. Different types of classification with and without priority to location can be performed. Likewise, location could also be used in clustering when there are no pre-classified training samples [8, 20]. Since customers who are likely to be close to each

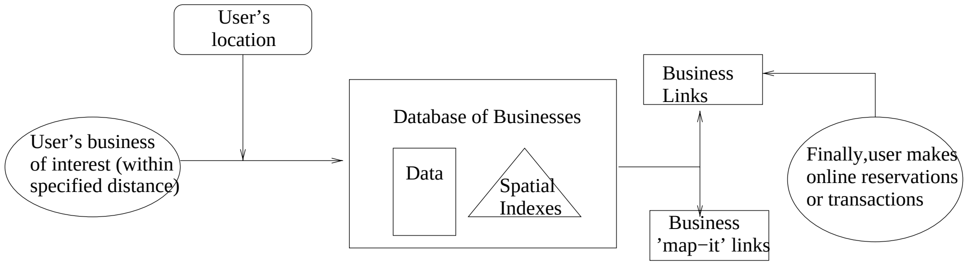

Figure 2: Using Location to Identify Businesses of interest In Mobile User's Region: Wireless service provider uses the customer's current location along with business type to identify interesting businesses (e.g. restaurants, movie theaters of interest) within a specified radius using its business database. User makes online reservations or transactions by going to the identified links.

other may exert influence on one another, [16, 2] describe new regression models that incorporate spatial autocorrelation in prediction.

Wireless service providers can provide location-based services such as nearest malls, theaters, restaurants, from the location of a mobile user. As illustrated in Figure 2, location-based services can be combined with e-business applications to serve mobile users. As part of the location-based services, the wireless provider maintains the database of e-businesses along with their locations. The data is indexed using spatial indexes such as Quadtrees or R-trees [14, 15, 19] to efficiently retrieve businesses within a specified distance from user-specified location. Partitioned local indexes could be constructed to restrict the spatial searches to businesses of the user-specified type. For example, all restaurants could be stored in a separate partition and all theaters as a separate partition. Having local-partitioned indexes for such partitions [12] would ensure only restaurants are searched when the user specifies the business type to be "restaurant". This approach ensures fast searching for business that satisfy user-constraints within the vicinity of a mobile user.

5 Conclusions

In this paper, we summarized the techniques for personalization of e-business sites. We described several approaches for such personalization using user's location as an input parameter. Such location-based personalization is quite useful in wireless applications where the location is the current mobile location of the user or in conventional static applications where the location refers to the address of the user.

References

[1] N. Beckmann, H. Kriegel, R. Schneider, and B. Seeger. The R* tree: An efficient and robust access method for points and rectangles. In *Proc. ACM SIGMOD Int. Conf. on Management of Data*, pages 322–331, 1990.

[2] S. Chawla, S. Shekhar, W. Wu, and U. Ozesmi. Extending data mining for spatial applications: A case study in predicting nest locations. In *ACM SIGMOD Workshop in Research issues in Data Mining and Knowledge Discovery*, Dallas, TX, 2000.

[3] C. Faloutsos, R. Barber, M. Flickner, J. Hafner, W. Niblack, D. Petkovic, and W. Equitz. Efficient and effective querying by image content. In *Journal of Intelligent Information Systems*, volume 3, pages 231–262, 1994.

[4] K. Instone. Information architecture and personalization. In *White paper: Argus Associates*, 2000.

[5] K. Koperski, J. Adhikary, and J. Han. Geominer: A system prototype for spatial data mining. In *Proc. ACM SIGMOD Int. Conf. on Management of Data*, Tucson, AZ, 1997.

[6] K. Koperski, J. Han, and J. Adhikary. Mining knowledge in geographical data. In *Commun. ACM*.

[7] K. Koperski, J. Han, and N. Stefanovic. An efficient two-step method for classification of spatial data. In *Symposium on Spatial Data Handling*, Vancouver, 1998.

[8] R. Ng and J. Han. Efficient and effective clustering methods for spatial data. In *Procóf the Int. Conf. on Very Large Data Bases*, Santiago, Chile, 1994.

[9] R. Ng and A. Sedighian. Evaluating multi-dimensional indexing structures for images transformed by principle component analysis. *Proc. of the SPIE*, 2670:50–61, 1994.

[10] I. Press. Ibm intelligent miner. In *IBM Documentation*, 2001.

[11] O. Press. Oracle personalization. In *Oracle Documentation*, 2001.

[12] O. Press. Partitioning in oracle. In *Oracle Documentation*, 2001.

[13] K. V. Ravi Kanth, D. Agrawal, Amr El Abbadi, and Ambuj K. Singh. Dimensionality reduction for similarity searching in dynamic databases. In *Proc. ACM SIGMOD Int. Conf. on Management of Data*, 1998.

[14] K. V. Ravi Kanth, Siva Ravada, J. Sharma, and J. Banerjee. Indexing medium-dimensionality data in oracle. In *Proc. ACM SIGMOD Int. Conf. on Management of Data*, 1999.

[15] H. Samet. Recent developments in linear quadtree-based geographic information systems. *Image and Vision Computing*, 5(3):187–197, Aug. 1987.

[16] S. Sekhar and Y. Huang. Discovering spatial co-location patterns: A summary of results. In *Symposium on Advanced Spatial and Temporal Databases*, Redondo Beach, CA, 2001.

[17] R. Shrikant and R. Agarwal. Mining generalized association rules. In *Procóf the Int. Conf. on Very Large Data Bases*, 1995.

[18] R. Shrikant and R. Agarwal. Mining quantitative association rules in large relational tables. In *Proc. ACM SIGMOD Int. Conf. on Management of Data*, 1996.

[19] F. Wang. Relational-linear quadtree approach for two-dimensional spatial representation and manipulation. *IEEE Trans. on Knowledge and Data Engineering*, 3(1):118–122, Mar. 1991.

[20] T. Zhang, R. Ramakrishnan, and M. Livny. BIRCH: A new data clustering algorithm and its applications. *Journal of Data Mining and Knowledge Discovery*, 1(2):141–182, 1995.