IBM Research Report

Cross-Channel Customer Mapping

Jayanta Basak
IBM Research Division
IBM India Research Lab
Block I, I.I.T. Campus, Hauz Khas
New Delhi - 110016. India.

Sunil Goyal
IBM Research Division
IBM India Research Lab
Block I, I.I.T. Campus, Hauz Khas
New Delhi - 110016. India.

IBM Research Division
Almaden - Austin - Beijing - Delhi - Haifa - T.J. Watson - Tokyo - Zurich

Abstract

In multi-channel setting (for example, web channel and storefront), it is often required to generate an integrated view of a customer across channels for making better CRM, marketing and merchandizing decisions. In order to generate integrated view, it is essential to uniquely identify a customer across channels. Due to various reasons, it is often not feasible to impose a unique identifier on a customer across channels. Moreover, a customer may not provide her true demographic information, which makes it even more difficult to track her. In the absence of a unique identifier and the demographic information, the behavioral profile of a customer can perhaps be used for tracking her across channels (cross-channel customer mapping). We define four channel-independent behavioral attributes namely, brand loyalty, price preference, responsiveness to promotion, and spending range. We call a behavioral attribute of a customer to be channel-independent if it does not change substantially with the channel characteristics in absence of a conscious effort from the customer. We defined the attributes in such a way that they reflect the statistical nature of a customer behavior, are easily computable, and can be incrementally updated over time. We then match these profile attribute values across channels to map the customers. We used two months Safeway transaction data; randomly divided it into two channels in different proportions, and computed these attribute values. We achieved significant accuracy in tracking the customers across channels, for example, more than 90% accuracy for the high valued customers.

1 Introduction

Merchants are increasingly offering products and services on multiple channels (e.g., mobile phones, PDAs, tele-sales and conventional channels such as stores, direct mail catalogs, and online retail sites) to allow broader reach and customer convenience. Although huge transaction data of the customers are available, the real task is to transform that customer data into useful intelligence that can help drive business decisions [13]. In order to do so, merchants need to create profile about how people shop and pay, how they behave over time, and how they react to different offers and prices. Using these profiles, merchants can then identify and set their priorities to influence customer behavior for better CRM and increased sales, profits, and wallet share.

In business analytics and data mining [6, 13, 17], the useful intelligence about the customers have been derived on a single channel so far. However, in order to be effective, for making better decision and taking actions, in a multi-channel setting a merchant needs to create an integrated view of each customer across channels. In order to generate an integrated view of a customer, it is essential to uniquely identify the customer across different channels – we identify this task as "cross-channel customer mapping".

One evident approach is to impose the same identity (for example, a customer-id) on a customer using different channels. However establishing the same identity may not be feasible for various reasons. First a customer may intentionally or unintentionally register in different channels with different identifiers. Secondly, a customer may not be aware of the fact that all these channels in question belong to the same merchant. This impression may be given in many cases as a merchant's multiple sales channels may operate with relative independence. Third, a merchant may have acquired a new sales-channel (such as through merger and acquisition) which has some common customer with the existing channels.

In general, customers can be described by two kinds of data – factual and transactional. The factual information consists of basic demographic data such as name, gender, birth date, address, income group. In general, the factual information is static [11]. The transactional data which is often dynamic, is composed of the purchases made by a customer and may include items purchased, date of purchases, amount, quantity, promotions, discounts used, discount values, payment type.

In the absence of any unique identifier, it is perhaps possible to track a customer across channels by using the factual and static information of the customer. However, often that is not feasible due to several reasons. First, it may not be possible to store the demographic data of all customers in the database. Second, even if it is possible, a customer may not update his demographic information, and third and the most compelling factor is that the customers may not be willing to reveal their true demographic information because of their perception about privacy and security. In the absence of any unique identifier and the factual information, a correspondence between the customers across channels can possibly be established by observing their behavioral profile from the transactional data.

In literature, various studies are made about the dynamic profiling of the customers [2, 3, 7, 9, 10, 11, 15, 16, 17, 18]. The profiling in the literature has various notions. For example, many of such studies are driven by the objective of extracting customer segments having certain homogeneous behavior. These segments are either extracted probabilistically (for example, modeling the transaction data as a linear combination of certain basis transactions) or performing certain clustering on the browsing or the purchase patterns. Various attempts are also made to create profiles by mining the data in the market basket for interesting rules [8, 12, 19]. Most of these profiling techniques bear the objective of obtaining better actionable items in a single channel (for example, it may be store-front or the web channel) [17]. However, in the multi-channel setting, the concept of profiling bear certain different notions in the sense that a homogeneous segment in one channel may not be homogeneous in the other channel. Secondly, the profiling should be such that they reveal the same characteristics of a customer – independent of the channels – so that the customer can be identified across channels. Creating profiles across channels has not been addressed in these kind of studies. In [1], the issues of cross channel marketing has been discussed where the effect of certain actions on one channel, on the other channel has been adaptively modeled by the technique of reinforcement learning. This technique addresses the gross behavior of the customers across the channels considering them as uniquely identifiable. However tracking an unknown customer across channels has not been addressed in the literature.

Cross-channel customer mapping involves steps of extracting channel-independent profile attribute information from customer behavior in different channels, and matching the channel-independent profile information across channels. A profile attribute can be taken to be channel independent if the techniques used for computing the value of the profile attribute do not depend on the channel characteristics. Conceptually, the customer does not consciously change her behavior across channels in respect of channel-independent profile attributes. For example, if a customer is loyal to some brand of a product (suggesting an underlying affinity of some kind with that brand) then she exhibits same loyalty across channels within certain reasonable duration. On the other hand, a customer may frequently visit a particular channel but another channel seldom. Therefore frequency of visit to a particular channel is, for example, not a channel-independent profile attribute.

In this paper, we identify a number of channel-independent profile attributes (behavioral characteristics) for identifying the customers. These attributes can be easily computed mainly from the transaction data and the product catalog. Moreover, these attributes can be incrementally updated. We then compare these profile vectors on two different channels. In comparing the profile attributes we show that simple Euclidian distance often do not give rise to the desired results and therefore we need to modify that. We tested the effectiveness of the proposed technique by experimenting with real-life transaction data and product catalog, and demonstrate that certain useful correspondence can actually be established with high accuracy.

2 Method

The overall method for establishing the correspondence between customers in the different channels, i.e., mapping customers from one channel to the other channel comprises of two major steps. First, certain channel-independent behavioral attributes are extracted for all customers on both channels, and then these attribute vectors are matched across channels.

2.1 Profile Extraction

The profile vector of a customer is comprised of her behavioral attribute values. The extraction of behavioral attributes is guided by three major requirements. First, the attributes should be independent of channel characteristics. Second, these attributes should be easy to compute from the data available to the merchant. We generally considered only the transaction data and the product catalog. Third, the attributes should be incrementally computable. Since the computation of the attribute values require huge data analysis, it may not be possible and practical to process the entire data every time to analyze the changing characteristics of the customers. Following we describe a set of such profile attributes.

1. **Loyalty:** Let L_{ip} denote the loyalty of a customer i in some product segment p. First, we define
$$L_{ij}(p) = \frac{X_{ij}(p)}{\sum_j X_{ij}(p)} \tag{1}$$
where $L_{ij}(p)$ represents the loyalty of the customer i to some brand j in the product segment p. $X_{ij}(p)$ is the amount spent by the customer on brand j in the product segment p in certain

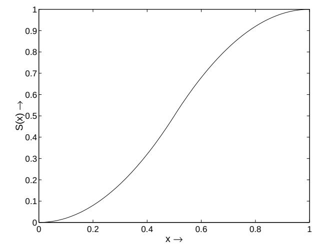


Figure 1: The nature of an S-function.

stipulated period of time. Then we define

$$L_{ip} = \max_{j}\{L_{ij}(p)\} \tag{2}$$

In other words,

$$L_{ip} = \frac{\max_j\{X_{ij}(p)\}}{\sum_j X_{ij}(p)} \tag{3}$$

In other words, we observe the consistency of a customer in purchasing the product of certain brand in a product category. If the customer is consistent then he is loyal in that product segment. Note that, we are considering the maximum value of the amount spent by the customer. In retail industry, the maximum or minimum value may badly affect the variables due to the presence of outliers. Instead of considering the simple maximum values, we can rescale the amount spent by the inter-quartile or inter-percentile range and then use the non-linearly scaled data to compute the loyalty in a product segment. However, in this article, we consider the task of nonlinear scaling of the data in the perspective of the data cleansing and outlier removal which can be performed separately before applying our technique. We therefore preserve the original amounts spent by the users. If we denote L_i as the overall loyalty of the customer i then

$$L_i = \frac{\sum_p L_{ip}}{M_i} \tag{4}$$

where M_i is the total number of product segments in which the customer i has transacted. In other words, we call a customer to be loyal if she is loyal in almost all the product segments that *he* **she** has transacted**in**.

It is interesting to note that the notion of loyalty can happen at different granularities. For example, we have defined $L_{ij}(p)$, L_{ip}, and L_i to denote the loyalty of a user i to a particular brand in product segment, in a particular product segment, and the over all product segments respectively. We can consider any one of them depending on the required accuracy and the available data in the store. In our experiments, we used the loyalty in a product segment i.e., L_{ip}, for matching the customers across channels. Note that, the notion of loyalty can have various other meanings also, e.g., loyalty to certain store, loyalty to a person and so on. The loyalty that we define here reflects the consistency of a customer in purchasing products of certain brands in various product segments. By 'product segment', we mean the product category which can be obtained from the product catalog. For example, 'red wine' can be a product category and it may consist of different brands.

The value of the loyalty (either L_{ip} or L_i) for a customer always lie in the range $[0, 1]$ irrespective of the amount spent. In order to compute L_{ip} incrementally, we need to maintain $\max_j\{X_{ij}(p)\}$ and $\sum_j X_{ij}(p)$ separately in the database tables. The value of L_{ip} can be updated with each transaction of the customer by a comparison and a summation.

It is possible to treat the customers who are very loyal to be similar as those who are ideally loyal. Similarly, the customers whose loyalties are very low can be treated as not loyal. Therefore, we can obtain such a characteristics by having a nonlinear transfer function on L_{ip} such that the brand loyalty can be defined as

$$\text{Brand Loyalty} = S(L_{ip}) \tag{5}$$

where $S(.)$ is a nonlinear function having nature as shown in Figure 1. However, it is not always necessary to have such a nonlinear transformation.

2. **Price Preference:** We observe the nature of a customer which reflects if the customer has affinity towards low priced items in a product segment or towards high priced items. In the marketing literature [14], in general, the price sensitivity of a customer or customer segment is measured. However, for the ease of computation, we computed the price preference of a customer.

   Let P_{ip} be the price preference of a customer i in the product segment p. We define

$$P_{ip} = \frac{x_i(p) - V_{min}(p)}{V_{max}(p) - V_{min}(p)} \tag{6}$$

   where $x_i(p)$ is the average per unit price paid by the customer for an item in the product segment p (note that it is different from $X_{ij}(p)$, although $x_i(p)$ can be derived from $X_{ij}(p)$). $V_{min}(p)$ is the minimum per unit price in the product segment p, and $V_{max}(p)$ is the maximum per unit price in the same product segment p. Irrespective of the amount spent by the customer, $P_{ip} \in [0, 1]$ where a low value of P_{ip} indicates that the customer i has affinity towards low priced items and a high value indicates a preference for high priced items. If we denote P_i as the overall price preference of the customer i then

$$P_i = \frac{\sum_p P_i p}{M_i} \tag{7}$$

   where M_i has the same meaning as above, i.e., the number of product segments in which the customer has transacted. Note that, the price preference can also be measured at different granularities as we described in the context of loyalty. The maximum and minimum amounts spent in a particular product segment can be noisy due to the presence of the data outliers. These data outliers can be handled by imposing certain non-linear scaling or the inter-percentile normalization of the data. However, here we do not include this data transformation in defining the attributes since the non-linear scaling can be performed before computing the customer attributes (we have stressed this point also in the context of loyalty).

   In order to distinguish between the characteristics for lower priced items and higher priced items, we can have a nonlinear transformation on P_i. Let $LowP$ and $HighP$ denote the preferences for lower priced items and the higher priced items. We can define

$$LowP_i = 1 - S(P_i) \tag{8}$$

   and

$$HighP_i = S(P_i) \tag{9}$$

   where $S(.)$ has the same characteristics as shown in Figure 1.

   In order to incrementally compute P_{ip}, we need to incrementally compute $x_i(p)$, since V_{min} and V_{max} are fixed (given that merchant did not update the prices). The value of $x_i(p)$ can be incrementally averaged. Even if the values of V_{max} and V_{min} change, it will be similarly reflected in the price preference of all customers, therefore the performance of profile matching will not be affected .

3. **Responsiveness to Promotion:** Let Ω_i be the set of coupons offered to customer i and ω_i be the set of coupons redeemed by the customer i. Let x_j be the absolute discount offered on coupon j. Then the responsiveness to high value offers (R_i) is defined by

$$R_i = S(\frac{v_i}{V_i}) \tag{10}$$

   where $V_i = \sum_{j \in \Omega_i} x_j$ is the total offered discount to customer i; and $v_i = \sum_{j \in \omega_i} x_j$ is the total amount of discount redemption by customer i, and $S(.)$ has the same characteristics as shown in Figure 1. The incremental computation is obvious by definition. One issue we must describe in the context of responsiveness to promotion. In general, it is very difficult to obtain the relevant data for computing the responsiveness to promotion from the retail stores. We therefore, consider this attribute as an optional one.

4. **Spending Range:** Let x_i be the average amount of purchases in certain standard unit by customer i over a period of time T. We define a variable y_i as

$$y_i = \frac{(x_i - \hat{x})}{\sigma(x)} \tag{11}$$

where $\hat{x}$ is the mean of x_i considering all the customers in the store, and $\sigma(x)$ is the standard deviation of the average amount of purchases over all customers. If Y denote the random variable corresponding to all customers' y_i, then the Spending Range of a customer is

$$B_i = P(Y = y_i) \tag{12}$$

where

$$P(Y = y_i) = \int_{-\infty}^{y_i} p(y)dy \tag{13}$$

Approximating $p(y)$ as a Gaussian $\mathcal{N}(0,1)$ distribution over all the customers,

$$B_i = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y_i} \exp(-\frac{u^2}{2})du \tag{14}$$

i.e.,

$$B_i = \frac{1}{2}\left[1 + erf(\frac{y_i}{\sqrt{2}})\right] \tag{15}$$

Note that, $B_i \in [0,1]$ where a low value of B_i indicates that the spending range of customer i is low and a high value indicates that the spending range is high. The attribute B_i is not very easily computable in the incremental mode. In order to incrementally compute B_i, we need to incrementally average x_i, $\hat{x}$, and $\sigma(x)$ from each transaction of each customer.

Note that, for the purpose of matching cutomers across channels, we define only four attributes (at different granularities). Various other attibutes can be defined and computed for such purpose, provided they fulfill basic conditions of 'simple to define' (more precisely, 'easy to compute') and 'channel-independence'. These attributes are also not necessarily complete for characterizing a customer. In retail, various other cutomer attributes are defined and used for different business objectives such as customer retention, targeting customer for increased revenue, cross-sell and up-sell, inventory clearance etc. For these different objectives various other customer attributes such as price sensitivity, loyalty program participation, affinity to specific product segments, etc. are used. The problem that we address i.e., cross-channel customer mapping does not deal with the similar set of end goals. Here we need to identify the same customer across channels. Therefore, it is unnecessary and useless to stick to the usual customer attributes driven by different business goals. Therefore we only consider attributes with invariant bahaviour across channels. For example, we did not consider the issues of price sensitivity at all. Rather, we considered the price preference which may not be directly useful in deciding product pricing for retailers, however, it is useful for identifying customers across channels. A similar explanation can be given for the attribute 'loyalty'. We quantify certain invariant characteristics of consistency about the customers and call it as 'loyalty' which is different from the notion of loyalty in the usual retail study.

2.2 Profile Matching

Once the profile attributes of the customers are extracted from the transaction data and the product catalog, these are matched across channels to establish the correspondence. Since our claim is that the extracted profile attributes are independent of the channel characteristics, they will be similar for the same customer across the channels. Note that, the profile attributes can represent the true characteristics of a customer provided these attributes are derived from large data sets (large volume of transaction) so that the true statistical properties are reflected therein. Therefore, it is expected that the customers who made large number of purchases can be better tracked across channels In other words, the customers who are of more value to the merchant (high-value customers) can be better tracked.

We derive the profile attribute values of each customer and then a customer is expressed by her profile vector. We then match the profile vector of a customer in one channel with the profile vectors of all customers in the other channel and retrieve the top K closest matches. The matching is performed by computing the Euclidian distance between the profile vectors. However, the Euclidian distance may not reflect the true dissimilarity between customers in different channels. For example, if a customer does not transact in a product segment then her loyalty in that product segment will be stored as zero. On the other hand, if she transacts only once then her loyalty in that segment will be one. Thus for customers who are not frequent buyers on both the channels, the Euclidian distance may be a misleading measure for dissimilarity. Instead of storing zero for the loyalty in a product segment, it can be stored as 'unknown' if a customer does not transact in that segment.

However, it is not possible to handle an 'unknown' value in the framework of Euclidian distance. We can at best compare only those loyalty values for which are known on both the channels. Comparing only the nonzero attribute values may also be misleading. For example, a customer i has purchased frequently in a product segment p and her loyalty L_{ip} is very high (close to unity). Another customer j does not like any brand in the product segment p at all, and L_{jp} is indeed zero. In that case, it makes sense to obtain a distance between L_{ip} and L_{jp} in a Euclidian framework. Therefore, when we compare the frequent customers on one channel with the less or more frequent customers on other channel, profile matching using Euclidian distance may make sense with the perception that the attribute values reflect their likes and dislikes. On the other hand, when we compare infrequent customers on one channels with the customers on the other channel, we should be concerned about the lack of information while using the Euclidian distance, and perhaps can use only non-zero attribute values for obtaining the dissimilarity. In the next section, we formally express this conceptualization with the real-life data. We consider both the cases, i.e., simple Euclidian distance and the restricted distance on the purchased product segments, and show the performance of the customer mapping in the next section.

3 Experiments

3.1 Experimental Protocol

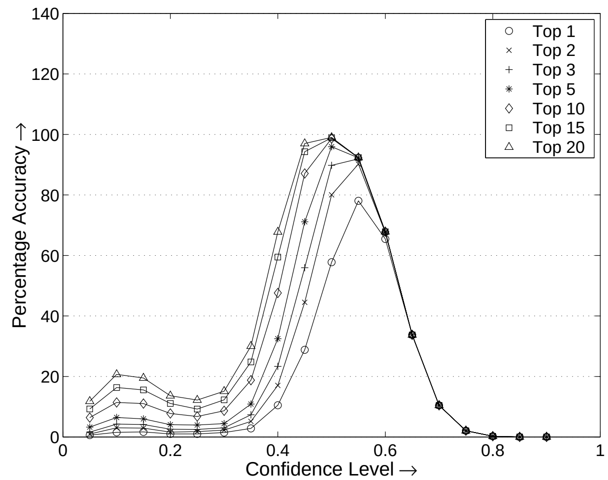
In order to evaluate the performance of the cross-channel customer mapping method, we need transaction data and product catalog in two different sales channels. However, in real-life it is very difficult to obtain such datasets. We therefore, used catalog and two-months transaction data (February and March, 1998) of the UK-based grocery store 'Safeway', and divided the transaction data randomly in two parts in order to emulate two sales channels. We divided the data in such a way that a complete transaction goes to one channel only. In other words, we did not divide a single transaction between two channels in order to reduce the effect of correlation between the channels. We divided the data in random proportions starting from $10 : 90$ to $90 : 10$, i.e., the number of purchases increases form 10% to 90% in one channel and decreases from 90% to 10% in another channel. We have performed this random division in order to investigate the performance of the customer mapping method from a 'low transaction volume' channel to a 'high transaction volume' channel and vice versa. We also observed the performance on different customer sets – from high value customers having large number of purchases to rather infrequent customers. Table 1 shows the five different customer sets that we considered having the number of items purchased as $100 - 199$, $200 - 299$, $300 - 399$, $400 - 499$ and more than 500. Note that, a transaction contains more than one such purchase. For example, if a customer purchases 20 items in a transaction on average then 200 purchases mean 10 such transaction. Secondly, the number of purchases as shown in Table 1 is the sum of purchases in both the channels. For example, the customer set having 500 purchases (Table 1) has approximately 250 purchases in the two months period in each channel when we divide the data in $50 : 50$ ration. They have approximately 50 purchases in one channel and 450 in the other when we divide the data into $10 : 90$ or $90 : 10$ ratio. We considered the attribute brand loyalty in a product segment (L_{ip}) only in order to map the customers across channels. This is done since the per unit price of the products were not available to us and therefore, the price preference could not be computed. Similarly, the responsiveness to marketing initiative could not be computed due to lack of the adequate information. We therefore demonstrate the effectiveness of our method based only on the loyalty in a product segment p, L_{ip}, of the each customer i. Computation of other profile attributes could have led to even better performance in tracking down the customers. Subsequently we demonstrate the performance of our method of customer mapping.

3.2 Experimental Results

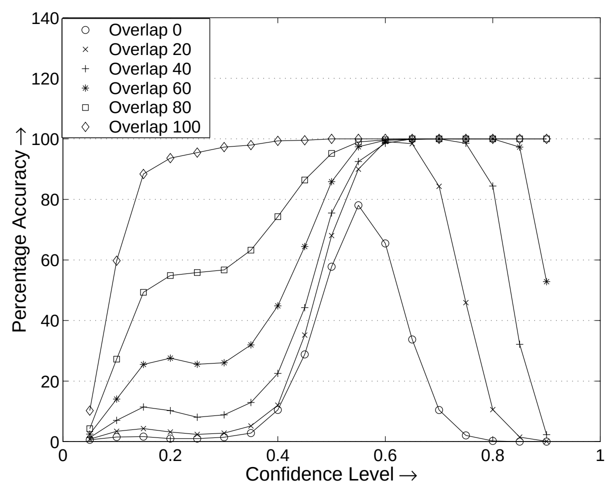
It has been mentioned earlier that the Euclidian distance does not reflect the true dissimilarity when the data is sparse, or rather the purchases are infrequent. We store the brand loyalty in a product segment as zero for a customer if he does not transact in that product segment. More precisely, the brand loyalty should be stored as 'unknown', although it is not clear how to deal with the attribute value 'unknown' in measuring the Euclidian distance. In order to address this problem, we can consider only those product segments in both the channels where the brand loyalty is nonzero. However, both the cases have certain advantages and disadvantages. In the following we illustrate that by defining a new dissimilarity measure.

Table 1: The different customer sets that we used for experimentation. The transaction volume represents the total number of purchases in both the channels. The brands and product categories for different customer sets are overlapping.

Transaction Volume	Number of Customers	Number of Categories	Number of Brands
100-199	7201	1274	2876
200 - 299	3222	1244	2658
300 - 399	1402	1166	2361
400 - 499	650	1104	2048
$\geq$500	440	1091	1996



(a)



(b)

Figure 2: The percentage accuracy in retrieving the correct customer. By percentage accuracy, we mean that how many times (in percentage) the customer in the query channel belongs in the top K customers retrieved from the other channel. The results reported here are an average over five different trials where in each trial we divide the transaction volume randomly in $50:50$ proportion into two channels. In this experimentation, we considered the customer set who have made 500+ total purchases in both the channels. The graph in (a) illustrates the performance for different K for the top K retrieved matches when there is no overlap between the transactions in two channels. The graph in (b) illustrates the performance of the top 1 match, i.e., exact match when there is overlap between the transactions. Note that, for 100% overlap, there are false negatives for the low confidence level since many customers have the same distance with the customer in query channel.

Characteristics of Dissimilarity: Let $[x_1, x_2, \cdots, x_n]$ represent the profile attribute values of some customer i in one channel, and $[y_1, y_2, \cdots, y_n]$ represent the profile attribute values of some customer j in another channel. We define

$$\Delta_i = |x_i - y_i| \tag{16}$$

and obtain a vector $[\Delta_1, \Delta_2, \cdots, \Delta_n]$. We sort these absolute differences in ascending order. Let the sorted differences be represented by $[\Delta^{(1)}, \Delta^{(2)}, \cdots, \Delta^{(n)}]$, and the corresponding profile attributes be $[x^{(1)}, x^{(2)}, \cdots, x^{(n)}]$ and $[y^{(1)}, y^{(2)}, \cdots, y^{(n)}]$. We define a dissimilarity between the two profile vectors as

$$d_k = \sqrt{\sum_{i=1}^{k} z(x^{(i)}) z(y^{(i)}) (\Delta^{(i)})^2} \tag{17}$$

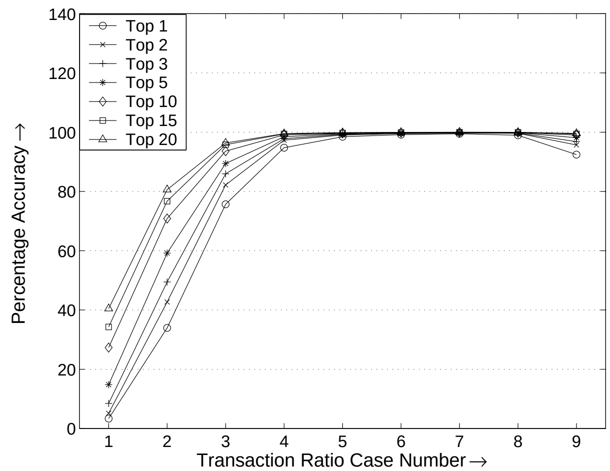
and a confidence value as

$$Conf_k = \frac{\sum_{i=1}^{k} z(x^{(i)}) z(y^{(i)})}{\sum_{i=1}^{n} z(x^{(i)})} \tag{18}$$

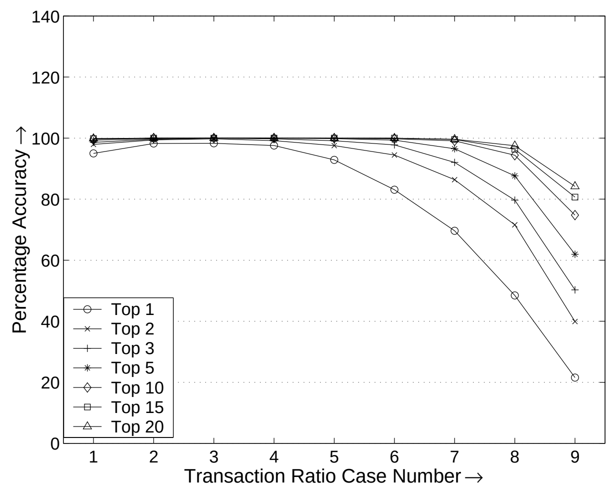
where $z(x) = 0$ if $x = 0$ and $z(x) = 1$ if $x > 0$. The distance d_k represents the difference considering the closest k profile attributes in two different channels. We considered Δ_i to be ∞ whenever either of the profile attribute values is zero i.e., we consider them as incomparable. The value $Conf_k$ represents the confidence level of the decision. For example, if there are 100 non-zero profile attribute values and we consider only ten of them then the confidence level is only 0.1. Note that, this dissimilarity is asymmetric in the sense that the same dissimilarity between two profile vector can occur with different confidence levels.

Given a customer in one channel (query channel), we find top K closest matches in the other channel using this dissimilarity measure. Note that, we conducted all the experiments considering only the profile attribute 'loyalty in a product segment', i.e., x_p corresponds to L_{ip}. We say that the retrieval is successful if the given customer belongs in the top K retrieved matches in the other channel. We obtain this retrieval performance using different confidence intervals $[c_1, c_2, \cdots, c_l]$. By confidence interval c_i, we mean that all matches obtains with dissimilarity measure with a confidence in the range $[(c_i + c_{i-1})/2, (c_i + c_{i+1})/2]$. Figure 2(a) illustrates the retrieval performance in top K closest matches for the customer set having more than 500 purchases (Table 1) in both the channels. The data is divided in the $50 : 50$ ratio. Figure 2(a) shows that the performance is very poor in the lower confidence values. This is due to the fact that with very low confidence any customer can be mapped with any other customer. Thus in the lower range of the confidence values we get false positives. As the confidence increases, the system becomes too prejudiced and starts finding the exact matches. Thus we get the false negatives. Note that in this graph, we consider the difference between two profile attribute values in two channels to be ∞ (very high) if either or both of them is zero. We also observed the performance by creating certain overlap between the transactions. Figure 2(b) illustrates the performance of top 1 match (i.e., exact match) for different overlaps among the transactions in the two channels. The graph shows the tendency that the performance becomes optimum for certain confidence level.

Performance for Various Ratios of Transaction Volumes on Two Channels: The graph in Figures 2 reveal two facts which are useful for matching the customer profiles in our next experimentation. First if we perform simple Euclidian distance then it might perform well when we map a customer from a high transaction volume (HTV) channel to a low transaction volume (LTV) channel. It is expected that in the LTV channel more profile attribute values will be zero (or rather 'unknown') as compared to the HTV channel. Therefore when we match a customer's profile on the HTV channel with the profiles on the LTV channel, we are expected to get less false negative (right side of the graph as shown in Figure 2(a)). However, if we perform the other way, i.e., we match a customer's profile on a LTV channel with the profiles in a HTV channel then we may get a large number of false positives. This is illustrated in Figure 3(a), where we demonstrate the performance based on Euclidian distance in matching customer profiles on two different channels. We observe a poor performance while matching, using the Euclidian distance, from a LTV channel to a HTV channel (say $10 : 90$ or $20 : 80$). The performance improves significantly when we match from a HTV channel to a LTV channel as shown in the Figure 3(a). All the curves shown in Figure 3(a) are for the customers who have made more than a total of 500 purchases in both channels. We modify the dissimilarity by considering the Euclidian distance of only those profile attribute values which are nonzero in the querying channel. In this case (from Figure 2), it is expected that we will get less false negatives when mapping from a LTV channel to a HTV channel, whereas we may get a rather large number of false positives when mapping from a HTV channel to a LTV channel. This trend is



(a)



(b)

Figure 3: The graphs in (a) illustrate the performance (in terms of percentage accuracy) in retrieving the correct customer in a set of top K customers from the other channel for various proportion of the transaction volume using the simple Euclidian distance. The transaction volume case numbers $1, 2, \cdots, 9$ represent proportions of the channel transaction volumes of $10:90$, $20:80$, $30:70$, $\cdots$, $80:20$, $90:10$ respectively. Note that the usage of Euclidian distance of the profile vectors causes a large number of false positives when matching from a LTV channel to a rather HTV channel. The graphs in (b) illustrate the same performance (i.e., retrieval performance in terms of percentage accuracy) using the modified dissimilarity which uses Euclidian distance for only the non-zeros profile attribute values in the query channel. Here we observe that usage of distance corresponding to only non-zero profile attributes causes a large number of false negatives while matching from a HTV channel to a LTV channel. All these graphs are the average results of five different trials where in each trial we divided the transaction data randomly into two channels in different proportions. The customers here considered are those who made a total of 500+ purchases in both the channels.

illustrated in Figure 3(b), where we demonstrate the performance of the system in retrieving similar customers based on the Euclidian distance of only those attributes which have non-zero values in the query channel.

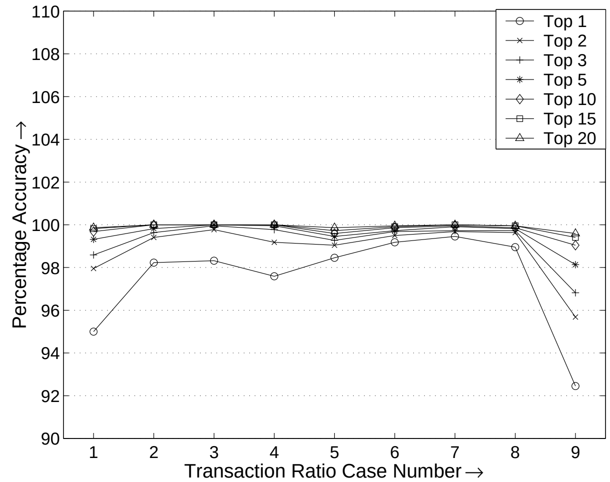


Figure 4: The retrieval performance in terms of percentage accuracy for different proportions of the transaction volume as in Figure 3. We merged the two different performances as in Figure 3. We used the modified dissimilarity (Euclidian distance considering only the non-zero attribute values of the query channel) while matching from a LTV channel to a HTV channel (for transaction volume ratio of $10:90$, $20:80$, $30:70$, and $40:60$), and the simple Euclidian distance while matching from a HTV channel to a LTV channel (for transaction volume ration of $50:50$ to $90:10$).All these graphs are the average results of five different trials where in each trial we divided the transaction data randomly into two channels in different proportions. The customers here considered are those who made a total of 500+ purchases in both the channels.

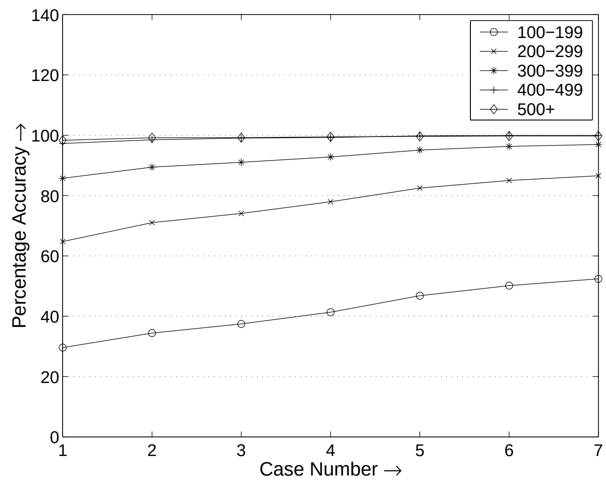
We therefore get the top K matches from a LTV channel to a HTV channel by using the Euclidian distance considering only the non-zero attributes. In order to get the top K matches from a HTV channel to a LTV channel, we use the simple Euclidian distance. Figure 4 illustrates the overall performance of the method for the high value customers who made more than a total of 500 purchases in both channels (Table 1). We observe that the method is able to map these customers across channels successfully with more than 90% accuracy.

Performance on Different Customer Sets: In our next experiment we demonstrate the performance for different customer sets. We considered the customer sets who have made a total purchases of $100-199$ items, $200-299$ items, $300-399$ items, $400-499$ items, and more than 500 items (Table 1) respectively in both the channels. We divided the transactions almost equally (i.e., $50:50$) in both channels. Figure 5 illustrates the performance of the system for different customer sets. We observe that the system performance degrades for the rather infrequent customers, however, it picks up very significantly for the average customers, and performs quite well for the high value customers.

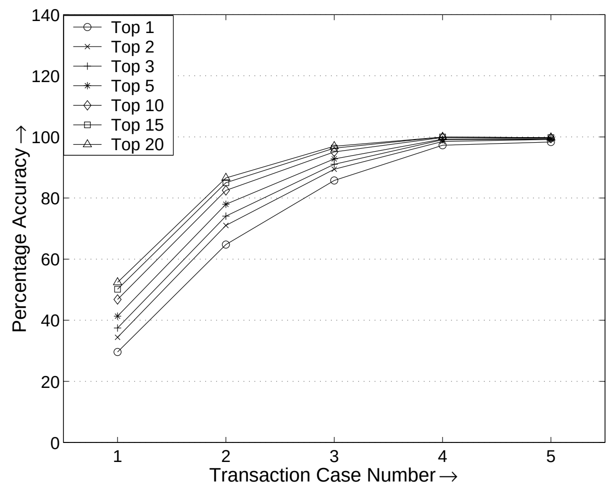
4 Discussion

We implemented the system of customer mapping form one channel to other channel. Figure 6 captures a snapshot of the overall system how it looks to the merchant. At present, the customer mapping is implemented for two channels, i.e., customers in one channel (query channel) are mapped *with* **to** the customers in the other channel. However, this can be extended to multiple channels for pair-wise mapping. We demonstrated the effectiveness of the system by experimenting with a real-life data set. We tested the system for different customer segments – from high value to rather infrequent customers. We observed that the method is able to map the high value customers very effectively. The performance for quite infrequent customers is not so good, although the performance for average customers is reasonably good. The method scales very well while mapping customers from high transaction volume channel to low transaction volume channel and vice versa.

There are several issues that can be further investigated. First of all, we used only one *kind* of profile attribute, i.e., brand loyalty (L_{ip}) for mapping a customer in the absence of requisite data.

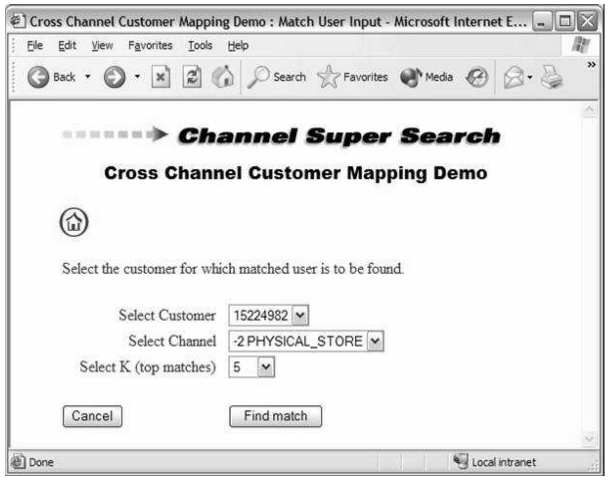


(a)

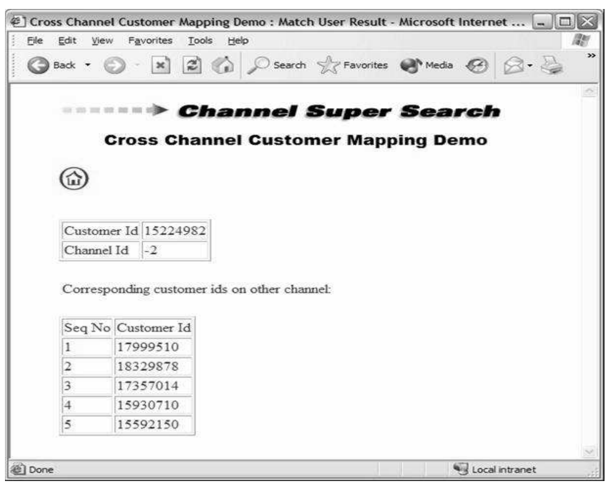


(b)

Figure 5: The graphs illustrate the performance in terms of percentage accuracy in retrieving the correct customer in top K matches for different customer sets. The performance is the average of five different trials with the transaction data randomly divided in 50 : 50 proportion in the two channels. The performance in all the cases are derived based on simple Euclidian distance. The curves in (a) illustrate the performance against the various cases which are top 1, top 2, top 3, top 5, top 10, top 15, and top 20 respectively and represented by 'Case Number' from 1 to 7 i.e., 'Case Number' 1 represents top 1 and 'Case Number' 7 represents top 20 respectively. The curves in (b) illustrate the same performance against different customer sets who made purchases of $100 - 199$, $200 - 299$, $300 - 399$, $400 - 499$, and 500+ items in both the channels. The different customer sets are represented by 'Transaction Case Number' from 1 to 5 such that 'Transaction Case Number' 1 represents the set of customers who made purchases of $100 - 199$ items in both the channels and 'Transaction Case Number' 5 represents the set of customers who made purchases of 500+ items in both the channels respectively.



(a)



(b)

Figure 6: A snapshot of the system in the operating condition. The figure in (a) demonstrates that a customer is selected and the value of K is specified so that system retrieves the top K closest matches. The figure in (b) shows the customers retrieved by our method.

Experimental results demonstrate certain potential in this approach even with this single attribute The system performance can possibly be further enhanced by using the other profile attributes as provided in Section 2. Secondly, we considered the same customer set in both the channels. It is not necessary that both the channels will have exactly the same customer set. However, the intent here is to demonstrate that the system is able to find out the same customer in the other channel if he exists. The present algorithm finds out the top K matches based on certain dissimilarity measure, although it does not incorporate any verification method as to if the retrieved customer is the same as in the query channel. In general, in the absence of any demographic information or identifier, it is difficult to incorporate one such verification system. This present model only tells the merchant about a set of customers in the other channel who are most similar to the one in the query channel. Third, the present model computes the profile attribute without addressing the temporal or seasonal variations. We computed these attributes by considering the average behavior over a certain period of time (here it is two months). The seasonal variations can possibly be captured by incorporating a sliding window over which the behavioral attributes are computed. Another major concern here is that we computed the attributes considering the product segments to be independent of each other. However, in real life certain association between the purchase patterns can affect the sales which can possibly be extracted by applying the data mining principles [4, 5, 6, 13]. Moreover, the purchase pattern in different sales channels may be affected by various other factors such as cross-sell and up-sell. One constraint in this method is that we considered the product catalogs of similar nature in different channels – although this may not be a big issue since most of the times the product catalogs in different channels (run by the same merchant) are very similar.

The extraction of behavioral profile of the customers has certain other advantages apart from the cross-channel customer mapping. Using the behavioral profile information, it is possible to make better decision and take actions such as targeting the valued customers. For example, a merchant may want to define an actionable customer segment and target this segment with the same promotions on a channel that were found effective (for example, in delivering sales and profit increases) on the other channel with a similar customer segment. However, to do so the merchant needs to be able to identify the similar customer segments across channels. Identifying similar customer segments across channels, if not identifying individual customers, is thus particularly desirable, and this can be effectively performed using our method.

5 Acknowledgments

We acknowledge Dr. Ravi Kothari for his keen interest and helpful suggestions in this work.

References

[1] N. Abe, N. Verma, C. Apte, and R. Schroko. Cross channel optimized marketing by reinforcement learning. In *Proc. Tenth ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD-2004), Industry/Government Track Poster, Seattle, Washington, USA*, pages 767–772, 2004.

[2] G. Adomavicius and A. Tuzhilin. User profiling in personalization applications through rule discovery and validation. In *Proc. Fifth ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD-1999), San Diego CA, USA*, pages 377–381, 1999.

[3] G. Adomavicius and A. Tuzhilin. Using data mining methods to build customer profiles. *IEEE Computer Magazine*, 34(2):74–82, 2001.

[4] R. Agrawal, H. Mannila, R. Srikant, H. Toivonen, and A. I. Verkamo. Fast discovery of association rules. In U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, editors, *Advances in Knowledge Discovery and Data Mining*, pages 307–328. AAAI Press, Menlo Park, CA, USA, 1996.

[5] R. Agrawal and G. Psaila. Active data mining. In *Proc. First Int. Conf. on Knowledge Discovery and Data Minining (KDD-1995), Montreal, Canada*, pages 3–8, 1995.

[6] S. Ansari, R. Kohavi, and Z. Zheng. Integrating e-commerce and data mining: Architecture and challenges. In *Proc. IEEE Int. Conf. on Data Mining (ICDM-2001), San Jose, CA, USA*, pages 27–34, 2001.

[7] C. Apte, E. Bibelnieks, R. Natarajan, E. Pednault, F. Tipu, D. Campbell, and B. Nelson. Segmentation-based modeling for advanced targeted marketing. In *Proc. Seventh ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD-2001), San Francisco, USA, August 26-29*, pages 408–413, 2001.

[8] C. Apte, B. Liu, E. P. D. Pednault, and P. Smyth. Evolving data mining into solutions for insights : Business application of data mining. *Communications of the ACM*, 45(8):49–53, 2002.

[9] I. V. Cadez, P. Smyth, and H. Mannila. Probabilistic modeling of transaction data with applications to profiling, visualization, and prediction. In *Proc. Seventh ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD-2001), San Francisco CA USA*, pages 37–46, 2001.

[10] P. B. Chou, E. Grossman, D. Gunopulos, and P. Kamesam. Identifying prospective customers. In *Proc. Seventh ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD-2000), Boston, USA*, pages 447–456, 2000.

[11] M. Eirinaki and M. Vazirgiannis. Web mining for web personalization. *ACM Trans. of Internet Technology*, 3(1):1–27, 2003.

[12] T. Fawcett and F. Provost. Combining data mining and machine learning for efficient user profiling. In *Proc. Second ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD-1996), Portland, USA*, pages 8–13, 1996.

[13] R. Kohavi, N. Rothleder, and E. Simoudis. Evolving data mining into solutions for insights : Emerging trends in business analytics. *Communications of the ACM*, 45(8):45–48, 2002.

[14] P. Kotler. *Marketing Management*. Prentice Hall, NJ, USA, 2002.

[15] H.-J. Lai, T.-P. Liang, and Y. C. Ku. Customized internet news services based on customer profiles. In *Proc. Fifth Int. Conf. Electronic Commerce (ICEC 2003), Pittsburgh, USA.*, pages 225–229, 2003.

[16] O. Nasraoui, H. Frigui, R. Krishnapuram, and A. Joshi. Extracting web user profiles using relational competitive fuzzy clustering. *Int. Journal Artificial Intell. Tools*, 9(4):509–526, 2000.

[17] J. Novo. *Drilling Down : Turning Customer Data into Profits with a Spreadsheet*. Jim Novo, St. Petersburg, USA, 2004.

[18] S. L. Pan and J. N. Lee. Using E-CRM for a unified view of the customer. *Communications of the ACM*, 46(4):95–99, 2003.

[19] M. Pazzani. A framework for collaborative, content-based and demographic filtering. *Artificial Intelligence Review*, 13(5-6):393–408, 1999.