Building Intelligent Shopping Assistants Using Individual Consumer Models

Chad Cumby, Andrew Fano, Rayid Ghani, Marko Krema
Accenture Technology Labs
161 N. Clark St
Chicago, IL, USA
chad.m.cumby,andrew.e.fano,rayid.ghani,marko.krema@accenture.com

ABSTRACT

This paper describes an Intelligent Shopping Assistant designed for a shopping cart mounted tablet PC that enables individual interactions with customers. We use machine learning algorithms to predict a shopping list for the customer's current trip and present this list on the device. As they navigate through the store, personalized promotions are presented using consumer models derived from loyalty card data for each inidvidual. In order for shopping assistant devices to be effective, we believe that they have to be powered by algorithms that are tuned for individual customers and can make accurate predictions about an individual's actions. We formally frame the shopping list prediction as a classification problem, describe the algorithms and methodology behind our system, and show that shopping list prediction can be done with high levels of accuracy, precision, and recall. Beyond the prediction of shopping lists we briefly introduce other aspects of the shopping assistant project, such as the use of consumer models to select appropriate promotional tactics, and the development of promotion planning simulation tools to enable retailers to plan personalized promotions delivered through such a shopping assistant.

Categories and Subject Descriptors: H.2.8 Database ManagementDatabase Applications[Data Mining]

General Terms: Algorithms, Economics, Experimentation.

Keywords: Retail applications, Machine learning, Classification.

1. INTRODUCTION

Supermarket shopping is an ideal environment to explore ubiquitous computing applications. Each week, millions of shoppers enter supermarkets in which they are immersed with tens of thousands of distinct product choices from which they will ultimately select a few dozen items, in about an hour or less.

We have built a prototype shopping assistant that aids supermarket shoppers and presents personalized promotions during the course of their visit. The prototype is designed for shopping cart mounted devices such as the one being piloted by Symbol Technologies. This device is essentially a Windows machine with a touch sensitive screen, a wireless barcode scanner, and built-in wireless access. The Symbol device is currently being field tested in stores with a self-scan and checkout application, in which shoppers scan items throughout the store. Our work builds upon this core capability by enabling personalized promotions. We obtained loyalty card data for a supermarket chain and used machine learning algorithms to build consumer models for each individual shopper.Our aim is to ground promotions in the context of an application that is viewed as a useful tool by the customer while providing business benefits to retailers and packaged goods companies.

A cart mounted device provides virtually continuous access to the customer. The challenge is how to use this access appropriately [1]. A supermarket visit may last an hour or more. An interaction that had the flavor of a rushed 30 second radio ad would be unbearable over the course of an entire shopping trip. Therefore two key issues to address are pacing – when we choose to use this access, and content – what we choose to communicate at these times. We understand that the customer is in control and can choose not to use our system if it is unpleasant, invasive, or useless.

Once the shopper is identified, through loyalty card or biometic identification, they are presented with a predicted shopping list for the current trip. It is not a set of promotional items or advertisements, but rather our best prediction of the items they *should* be buying today based on their past behavior. The list is meant to be useful to the shopper as well as benefit the retailer by reclaiming forgotten purchases. In our preliminary analysis approximately 11% of purchases are forgotten by shoppers, based on examining the deviation of replenishment intervals.

The shopping list serves as a key anchor for subsequent interactions during the course of the shopping trip. In our application, as the shopper travels throughout the various aisles of the store they are presented with items from the predicted shopping list available in the current aisle. In addition to the items from the shopping list, lower probability items that the shopper has purchased in the past are included in a list titled *Also of Interest*. To avoid overwhelming the user we only alter the list presented twice per aisle and while we haven't settled on a maximum number of items to show at any given time, we intend to keep it below 12.

The fact that a product category is on a shopping list and that the user is in the appropriate aisle provides the opportunity to present a promotion for that category. The next problem is determining what kind of promotion to provide, if any. We are addressing this issue by relying on in-

[1]Some examples of systems addressing this interaction on a non-individual basis can be seen in [1, 2, 3]

Copyright is held by the author/owner.
IUI'05, January 9–12, 2005, San Diego, California, USA.
ACM 1-58113-894-6/05/0001.

dividual consumer models consisting of over 100 attributes that characterize their purchasing behaviors. These models cover such areas as price sensitivity, promotional sensitivity, brand loyalty, basket size variability, among others, and are calculated for each shopper and product. In addition to predicting shopping lists we use these models to select between a variety of promotional tactics, such as larger pack sizes to increase basket size, brand extensions, and complement offers – to name a few.

Because our promotions are presented as a function of the predicted shopping list, the accuracy of these predictions is crucial. We therefore dedicate the main part of this paper to discussing how we predict the shopping list and postpone detailed discussion of other aspects of the behavioral models and promotion selection tactics for another publication.

2. MOTIVATION FOR INDIVIDUAL CONSUMER MODELS

Loyalty card programs at many grocery chains have resulted in the capture of millions of transactions and purchases directly associated with the customers making them. Traditionally, most of the data mining work using retail transaction data has focused on approaches that use clustering or segmentation strategies. Each customer is "profiled" based on other "similar" customers and placed in one (or more) clusters. This is usually done to overcome the data sparseness problem and results in systems that are able to overcome the variance in the shopping behaviors of individual customers, while losing precision on any one customer. We believe that given the massive amounts of data being captured, and the relative high shopping frequency of a grocery store customer, we can develop individual consumer models that are based on only a single customer's historical data. Our hypothesis is that by utilizing the detailed transaction records to build separate classifiers for every unique customer, we can improve on the performance of clustering and segmentation approaches and provide a more personalized experience to the customer.

3. SHOPPING LIST PREDICTION

This section explains the methodology behind our shopping list predictor, as well as the evaluation criteria we use to judge its success. Due to space limitations, we wll not include details in this paper. A detailed, in-depth description of the shopping list prediction is in [4].

We present some baseline methods to predict customer shopping lists that we can hopefully improve on using machine learning techniques. These include a random baseline (every category that the customer has purchased before has a chance proportional to its purchase frequency of being included in the shopping list), a Same as Last Trip baseline, and a top N baseline (N most bought products by that customer). We frame the problem of predicting the overall assortment of categories purchased as a classification problem. Each class can be thought of as a customer and product category pair. If our data set represents a customer set C and an average of q categories bought by each customer, we construct $|C| \times q$ classes y (and as many binary classifiers). For each of these classes y_i, a classifier is trained in the supervised learning paradigm to predict whether that category will be bought by that customer in that particular transaction.

	Recall	**Prec**	**F-Measure**	**Accuracy**
random	.20	.19	.20	.65
sameas	.25	.29	.27	.70
top-10	.41	.33	.37	.65
Perceptron	.40	.27	.32	.66
Winnow	.17	.38	.24	.79
C4.5	.25	.28	.26	.70
Hybrid-Per	.60	.27	.37	.55
Hybrid-Win	.44	.32	.37	.64
Hybrid-C4.5	.48	.34	.40	.62

Table 1: Transaction averaged results.

We experimented with two kinds of machine learning methods to perform this task – Decision trees (specifically, C4.5), and several linear methods (Perceptron, Winnow, and Naive Bayes) to learn each class . These linear methods offer several advantages in a real-world setting, most notably the quick evaluation of generated hypotheses and their ability to be trained in an on-line fashion.

In each case, a feature extraction step preceded the learning phase. Information about each transaction t is encoded as a vector in $\Re^n$. For each transaction, we include properties of the current visit to the store, as well as information about the local history before that date in terms of data about the previous 4 transactions.

We also explored hybrid methods by combining the *top n* baseline classifier with the various learned classifiers. If the *top n* predictor (for given n) is positive for a given class, then we predict positive, otherwise we predict according to the output of a given learned predictor.

3.1 Evaluation

We use *recall*, *precision*, *accuracy* and *f-measure* as our evaluation metrics. Typically, these metrics are aggregated in several ways. *Microaveraged* results are obtained by aggregating the test examples from all classes together and evaluating each metric over the entire set. The alternative is to *macroaverage*, in which case we evaluate each metric over each class separately, and then average the results over all classes. The transactional nature of the purchase datasource allows us to aggregate all examples associated with a single customer, obtain results for the above metrics for each set, and average them. We call this *customer averaging.* We also aggregate all the examples from each transaction, calculate each metric, and average the results over all transactions, which we call *transaction averaging.*

4. EXPERIMENTS & RESULTS

The dataset used contains transaction based purchase data for over 150,000 customers from a grocery store collected over two years. This population was sampled to produce a dataset of 2200 customers with 146,000 associated transactions. Results are shown in Table 1 for each approach, broken down by the *transaction* and *customer* averaging methods mentioned in the previous section.

Many of the results are promising in the context of predicting shopping lists for a large number of grocery customers. In terms of providing useful suggestions, we would like to obtain results that cover most of the items in a customer's potential shopping list (high recall) while not overloading the customer with a long list of non-relevant items

	recaptured
top10	10620
Perceptron	20244
Winnow	5251
C4.5	9134
Hybrid-Per	23489
Hybrid-Win	12270
Hybrid-C4.5	15405

Table 2: Number of forgotten purchases recaptured.

(precision). Our results show that it is difficult to accurately predict over 50% of the bought categories with a reasonable level of precision. Only the hybrid method combining the *top 10* classifier with the Perceptron based classifier achieves this high level of recall.

In general each hybrid method performs much better than all the other methods. Each obtains a significantly higher level of recall than its individual component classifiers, with comparable levels of precision.

Due to the wildly imbalanced training set sizes across classes both within and without customer groupings, many classes may contain very few positive examples. The baseline *top-10* classifier gives us a basic level of recall across all classes regardless of the training set size, while the learned classifiers would very rarely produce true positives for these classes. For classes with large training set sizes, using the learned classifiers gives us an advantage in terms of precision and accuracy. A distinctive feature of the data source we use is its high degree of systematic (non-random) noise due to customers forgetting to buy items they intended to buy. Based on the assumptions made about the distribution of forgotten purchases in the dataset, we can estimate the degree to which classifiers used in our experiments are robust to the label noise. For example, several of the algorithms exhibit enhanced precision when labels for instances of forgetting are manually flipped to become positive, while the random baseline technique performs the same. While the number of true positives do increase, not all the added positive examples are classified correctly, so in some cases the overall recall decreases or remains constant. But in Table 2 we show the number of added positive examples "recaptured" by the different classification algorithms, suggesting a measure of their relative robustness. The total number of examples for which we flip labels from negative to positive throughout all test sets in this case is 47916. This number represents a relative upper bound for the amount of purchases we can recapture given our assumptions.

5. CONCLUSION & FUTURE WORK

Our results thus far suggest that anchoring the customer interaction of a shopping assistant around a predicted shopping list is a viable approach. Our ongoing work is focused along three dimensions: Improved consumer models and prediction, promotion selection and deployment, and promotion planning.

The prediction of a shopping list is a prediction of one kind of behavior – the decision to buy an item within a particular product category. Our interest, however, lies beyond this one behavior. Much of our current work lies in deriving consumer models that enable selection of promotions that a shopper is likely to use and that a retailer is likely to benefit from, if they are, in fact, used. We are therefore working to incorporate a wide variety of behavioral attributes that characterize different shopping behaviors, including those that may be exhibited over multiple trips. One such example is pantry loading. An aggressive pantry loader will take advantage of a sale to stockpile a product, and then forego purchases at regular prices on several subsequent visits with no net increase in consumption, resulting in little or no benefit for the retailer or packaged goods company. Detecting and including such an attribute in a consumer model enables retailers to present promotions with benefits to both the retailer and consumer, rather than including those that may hurt the retailer or be of no interest to the consumer.

Beyond the enabling of personal promotions, a commercially viable shopping assistant would likely include additional features such as a map of the store, product locator, child entertainment, recipes, invocation of in-store services such as the deli, among others. While such features would be implemented in any final, deployed system, they are beyond the scope of our current research.

The personal promotion approach described here arguably enables promotions to be sold by retailers to packaged goods companies in a very different way. Typically retailers sell tactics such as endcaps (those displays at the end of aisles) or a space in the weekend circular. However, the availability of a channel to individual customers coupled with the ability to measure individual responses enables retailers to sell promotional results. We are therefore working on promotion design tools that enable retailers to simulate the deployment of personalized promotions to particular populations over a given period of time. This will allow retailers to make reasonable estimates for what can be reasonably be achieved with a promotion and at what cost. Given that our current consumer models are derived from data that have not included promotions of the kind described here, our promotion planning simulations include qualitative estimates of factors such as conversion rates for different promotional tactics. We are currently exploring opportunities to field test a version of the shopping assistant that will allow us to quantify these attributes more rigorously. The intent of the simulations at this point is to demonstrate a different approach for selling promotions.

In the end ubiquitous computing infrastructure and applications such as cart mounted tablets and shopping assistants will be paid for by business. Shoppers will expect to use these applications for free. Conversely, consumers will always have the choice not to use an application if it is perceived as annoying, invasive, or useless. In our research we have therefore strived to pay as much attention to ensuring ways to achieve the business value that will pay for and enable intelligent shopping assistants, as to the value provided to the consumer.

6. REFERENCES

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

[2] R. Bellamy, J. Brezin, W. Kellogg, and J. Richards. Designing an e-grocery application for a palm computer: Usability and interface issues. *IEEE Communications*, 8(4), 2001.

[3] E. Newcomb, T. Pashley, and J. Stasko. Mobile computing in the retail arena. In *CHI2003*, pages 337–344. ACM Press, 2003.

[4] C. Cumby, A. Fano, R. Ghani, and M. Krema. Predicting Customer Shopping Lists from Point-of-Sale Purchase Data In *KDD2004*, pages 402–409. ACM Press, 2004.