

Memory Zones for Online Supermarket Shopping

Martin Halvey and Mark T Keane

Abstract—Problems occur for users when they attempt to navigate or purchase items from a virtual shopping environment that has to correlate in the physical world. These problems arise in part, because online shoppers do not have the benefit of external memory provided by the physical world. The memory zones idea attempts to solve these problems by providing an online parallel to this external memory. The system reported uses a profile of previous purchases and a representation of the physical environment of the actual shop to recommend items to shoppers, that they may have otherwise forgotten because of the lack of suitable external memory

Index Terms—Online Shopping, Personalization, Recommenders, Spatial Representation.

I. INTRODUCTION

Online shopping for groceries and other household goods from your local supermarket is now becoming more and more common (e.g. [1], [2], [3]). Though many online systems implicitly assume that only operate in the online, virtual world, most shoppers continue to alternate between their virtual-world and real-world supermarkets. We believe that this state of affairs will continue for the foreseeable future. The main proposition of this paper is that the dual-world shopping experience can be leveraged to improve the usability of online shopping systems through the use of the idea of memory zones.

Online supermarket shopping for food and household goods is not as easy as it should be and manifests key usability problems. The main difficulties arise in (i) finding the items to buy in hierarchies and lists by tedious navigation or imprecise keyword searches and (ii) remembering all of the items that one wants to buy. In the real world most people visit the supermarket with a partial list of the items they want to buy and then wander the aisles picking up additional items based on what they see on the shelves. In other words, they use a mixture of their own memory, the external memory of their shopping list and the external memory of the actual store. In the online world, most shoppers may also have a partial list of required items (or provided lists from previous orders) but they do not

have the external memory of the store. This mismatch between the virtual and actual world often results in many items being forgotten, requiring ironically, follow-up visits to the real-world shops. Our memory zones idea aims to provide an online parallel to the external memory of the store's shelves to solve this problem.

Of course, many online shopping sites (e.g. www.buy4now.ie/superquinn) recognize these problems and try to solve them in a variety of ways. A typical solution is to provide “forgetting lists” of frequently bought items. However, this is only a partial solution. Apart from requiring extra navigation on the part of the user, these lists often omit the very items that people forget. Typically the external memory of the store reminds people of those items they buy irregularly (e.g., light bulbs, oven cleaner, rubber gloves); the items that are bought once a month or every second week rather than every week. Such items do not find their way onto forgetting lists because they are, by definition, infrequent purchases. To put it another way, forgetting lists probably remind people of things they remember anyway. Neither could forgetting lists be simply amended with these infrequent items, as they would soon become unwieldy listings of almost every item ever purchased by the user.

Our solution on the forgetting problem is quite different and, we believe, makes more sense from a cognitive perspective [4]. It is simply to try to reinstate some aspects of the actual world in the online world, specifically to model the layout of the actual store used normally by the shopper. Then using recommender techniques that analyse previous purchases we attempt to simulate the external memory experience by proposing near-by, possibly forgotten purchases as the shopper chooses items in the online store. This is the essence of the memory zones idea. In the following sections, we first outline some of the background literature before presenting the Memory Zones System and some empirical data on optimal parameters for the system.

II. BACKGROUND AND RELATED WORK

The general problem of information overload in the online shopping domain has presented researchers with many severe challenges to improve usability. A large body of work has concentrated on developing new interface technologies to support the users task (e.g., [5], [6]). However, this work is more likely to deliver medium-term to long-term success given the current delivery bottlenecks in the Web. More successful short-term solutions have been developed using recommender systems that work within current bandwidth limitations (e.g., [7], [8]). Given that we

This material is based on works supported by Science Foundation Ireland under Grant No. 03/IN.3/1361 to the second author.

Martin Halvey is with the Adaptive Information Cluster, Smart Media Institute, Department of Computer Science, UCD, Belfield, Dublin, Ireland. (telephone: +353-01-7162930, e-mail: martin.halvey@ucd.ie).

Mark Keane is with the Adaptive Information Cluster, Smart Media Institute, Department of Computer Science, UCD, Belfield, Dublin, Ireland. (telephone: +353-01-7062470, e-mail: mark.keane@ucd.ie).

seem to be stuck with list-based shopping systems for the near future, it makes sense to develop personalization methods for optimising the information that they convey. Typically, web based recommender systems suggest relevant items to users and/or filter out irrelevant items based on some analysis of the user's buying history and/or the history of other users [see e.g., 9, 10, 11]. Broadly speaking these systems can be divided into model- and memory-based systems. In model-based systems, purchasing information is analysed to characterise commonly occurring patterns, often captured in association rules (e.g., [12], [13]). For example, an association rule might capture the fact that 98% of customers that get new tyres on their car also get some other maintenance work done to their car at the same time (see [4]). These methods have been successfully applied to the basket data (i.e., transaction dates and volumes) that form the basis of supermarket shopping, though there can be significant overheads in maintaining and creating these association rules (e.g., AIS [13], SETM [14], Apriori [13] and AprioriTid [13]).

In memory-based systems, such as those using collaborative filtering, users with similar interests are grouped together and this information is used to leverage recommendations. If two users have consistently liked or disliked the same TV programs then the probability is that they will also like or dislike other similar TV programs. There are many successful examples of such systems (Ringo [15], Grouplens[16], PTV[7], [17] and the Tapestry System[9]). Memory-based recommenders can also use content-based filtering where objects are given a fixed set of attributes, and items are recommended to users based on a correlation between the users search and the attributes of an object (see e.g., Libra [18]).

III. THE MEMORY ZONES RECOMMENDER

Our recommender system, the Memory Zones Recommender, tries to reinstate the some aspects of the external memory of the store in the online context. It does this by exploiting a spatial model of the store and an analysis of the shopper's previous purchases. As such the three main components of the system: a component for (i) modeling the store's products and layout, (ii) modeling user's purchasing history, and (iii) making recommendations. All of these components act together to provide recommendations to the user.

A. Modeling store's products and layout

Most supermarkets have a definite organization of products into sections that are often structured to influence the goods purchased by shoppers [19]. Items can be organized by type and spatially, but these organizational schemes may not always overlap. For example, different teas would tend to be located together and tinned foods are often together. But, tinned tomatoes are often in the dried pasta section because they are an essential ingredient in Italian dishes. To capture store information the Memory Zones System has two different data structures: a type hierarchy of

products and a layout graph of product locations. The *Type Hierarchy of Products* is a content-based hierarchy although it is also partly captures location information (akin to those used to aid search in current online systems). Our hierarchy has as a root node corresponding to a major section of the shop (e.g., dairy, health/beauty), with child nodes that are sub-sections in the shop (e.g., dairy might be split up into milk, yogurt, cream). The leaf nodes in the hierarchy are the actual groceries that are on sale in the store (see Fig 1).

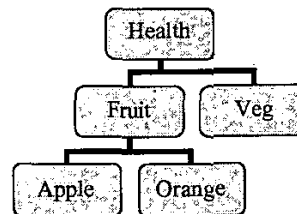


Fig 1. Example of a portion the product hierarchy.

The *Layout Graph* captures more fine-grained information about the location of products in the shop. In the actual store, sections are organised spatially in relation to each other and are connected by aisles. In the Memory Zones System, the layout of the store is represented as a weighted graph, with the nodes of the graph being the sections of the shop and the edges of the graph being the aisles in the shop.

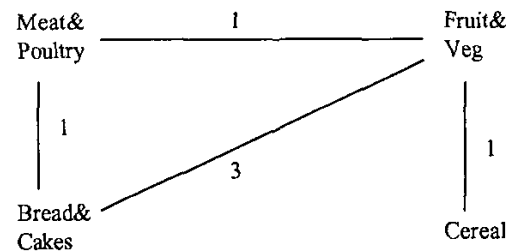


Fig 2: Portion of the Layout Graph

The weights of the edges represent the distance between two sections in the shop. Furthermore, the nodes of the graph map onto the root nodes of the product hierarchy. One way of viewing the situation is that each node of the graph is a tree, although they are still completely different hierarchies. By combining the knowledge of what is in a section with that of the graph representation of the shop, the relative distances between any two groceries can easily be determined.

B. Modeling a Shopper's Previous Purchases

The second main component of the Memory Zones System deals with a customer's previous purchases. Each customer in the system has a list of all of their previous purchases associated with them. These previous purchases include information about what was purchased and when.

This information is used to create a memory zone for each customer that informs the recommendations made. Each grocery in the system is assigned a BuyFrequency with respect to each customer. *BuyFrequency* is the number

of times that the user has purchased that particular grocery from the shop, divided by the total number of purchasing occasions carried out by the customer (i.e., online orders or shop visits). Each product is given a PurchaseProbability based on the BuyFrequency and time elapsed since that product was last purchased according to:

$$\text{PurchaseProbability}(x) = \text{BuyFrequency}(x) * \text{TimeUnit} \quad (1)$$

where a TimeUnit is the number of time steps in weeks since the last purchase of the product, it is recomputed each time a user visits the online store. PurchaseProbability is assigned to every item purchased by a particular user, forming part of that user's profile.

C. Making Recommendations

The Memory Zones System makes recommendations taking the user's purchasing history and the layout of the store into account. Specifically, an item is recommended based on its PurchasingProbability and its proximity to the last item bought. For example, if a user is buying toothpaste, that user is more likely to be recommended mouthwash rather than a barbeque, as mouthwash is close to toothpaste in the shop than a barbeque. The recommendation function is as follows:

$$\text{RecommendationWeight}(y) = \frac{\text{PurchaseProbability}(y)}{\text{Distance}(x,y)} \quad (2)$$

Where y is an item in the store, x is the item that is currently being purchased and the PurchaseProbability is the same as has been described in the previous section and Distance(x,y) is the distance between the two items in the layout representation. In this way, we try to simulate the real-world experience of reaching for one item off the shelf and noticing another that we sometimes buy and realizing that we need it though it was not on our shopping list. Apart from this form of recommendation the Memory Zone System also makes special offer recommendations, uses association rules for spotting paired purchases and uses collaborative filtering to capture customer likes/dislikes.

Special Offers. Apart from simulating an external memory of the store, this recommendation function can also be used to make special offer recommendations to shoppers. The system maintains a special offer set of groceries, and if the customer is about to buy an item that is in the same section as an item in the special offer set, then the special offer can be recommended to the user. For example if a user was due to buy Colgate toothpaste and the special offer set contained MacLean's toothpaste, then MacLean's will be recommended to the user.

Paired Purchases. The system also uses association rules to identify paired purchases. For example, if a user previously purchased a barbeque and also around the same time purchased a fire extinguisher then in the future if that user buys a barbeque they may also again require a fire extinguisher. In the Memory Zones System, if two items have the same frequency and were last purchased on the same date, then according to the system they will both be purchased again on the same date in the future. So the system will form pairs of these items and if one item is

purchased, then it is assumed that the other will also be purchased on the same date. So the second item of the pair will be recommended to the user.

Likes & Dislikes. Collaborative filtering is also implemented in the system to identify users' likes and dislikes. Customers are compared on their previous purchases and those that have bought similar items in the past are grouped together. So if user A and user B are nearest neighbours, and user A has an item that they buy frequently that user B has never purchased, then that particular item will be recommended to user B. In this way items that a user has not previously purchased but that similar users often purchase can be recommended.

IV. CALIBRATING PURCHASES

In the previous section, we outlined the basic way in which recommendation works in the Memory Zones System. However, it is clear that such recommendations could get very annoying to a user, if every other purchase that you made in a certain part of the store was suggested to you every time to buy a product. To be effective the system needs to recommend items that the person may really have forgotten. In system terms, we need to calibrate it to find the optimal PurchaseProbability threshold at which an item should be recommended.

A. Study 1: Optimal Thresholds for Memory Zones

We determined this threshold by performing an empirical test on a database of purchases from two different customers at the same store. In essence, we tried to find the threshold value that best predicted the items bought on a fifth purchasing session, using user profiles built up over four previous sessions.

1) Experimental Set-up

We gathered records of five purchasing sessions for two customers at the same store. The average number of purchases per session was 53, with a minimum of 40 and a maximum of 63. All five sessions for one of the customers were from visits to the actual store and all five sessions for second customer were based on online purchasing sessions. The layout of the actual store (i.e., Superquinn's Supermarket at Sutton Cross, Dublin, Ireland) was encoded in a layout graph for all of the items purchased by these customers, with a few additional items added. The first four sessions were used to compute user profiles for each customer. The appropriate PurchaseProbabilities for each item bought were computed. We then varied the threshold value for adding groceries to the customer's memory zone in a range between 0.25 and 1.5 and noted the proportion of items from the fifth, unprocessed session that were correctly predicted at each threshold level.

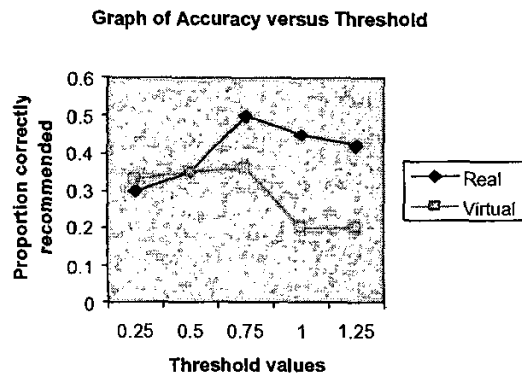


Fig 3. Graph illustrating the results of the first experiment.

2) Results

The results showed that the system properly predicted between 30%-50% of the items in the fifth shopping session based on the analysis of four previous sessions by the customer. On the whole, the proportion of correct predictions was higher for customer purchasing in the real world than for the customer purchasing on-line. However, the optimal threshold for both customers was around a PurchasingProbability of 0.75. At lower threshold values almost all of the previous purchases for each customer are added to that customer's memory zone. At the higher threshold values, frequently bought groceries are excluded from the customer's memory zone.

3) Discussion

The main positive result from this experiment is the finding that there is a clear threshold beyond which the prediction accuracy of to-be-bought decreases. Furthermore, from a processing overhead point of view, it is good that such levels of prediction can be achieved on the basis of a small sample of sessions (4). A larger sample of sessions might well deliver better predictive accuracy but would also attract an unwelcome processing overhead.

At first sight, the predictions based on the real-world shopper versus the online one look worrying. However, we believe that this merely reflects the greater success of real world shopping versus on-line shopping. The sample of items purchased on-line is a subset of the items the customer wanted to buy and, as such, is a poorer reflection of what the customer normally buys. The customer in question confirmed this to us, saying that the online orders we used omitted many items that had to be bought either the following week or at the local shop. The more complete lists of the real-world shopper were a better reflection of what is normally bought. There is an important lesson here; namely, that it would be better to mine sessions when the user visited the shop for recommendations than the online sessions. Of course, this exactly illustrates the point we want to make, that the real-world experience should be used to leverage the online one.

B. Study2: Improving Accuracy in the Virtual

In the second experiment, we followed our conjecture that the predictive inaccuracy for the online customer was partly the result of the poor online shopping experience. We suggested that the lower predictive accuracy may have been due to the fact that many items that the customer intended to buy were forgotten. If this were true then if we

include purchases from a following week, the sixth session, then the predictive accuracy should improve. To put it another way, if we combine what was purchased in the fifth and sixth sessions of the online buyer we may then have a sample that better corresponds to the fifth real-world session of the other customer.

1) Experimental Set-up

The set-up was identical to that of the first study except that we noted the predictive accuracy of the different threshold values for the combined fifth and sixth sessions of the online shopper.

Graph of Accuracy versus Threshold

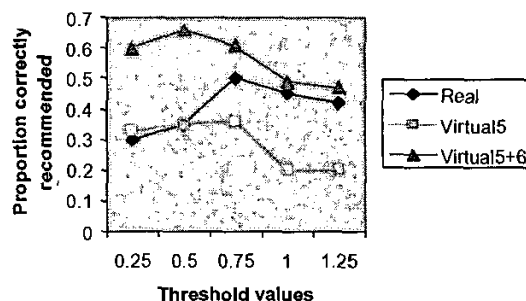


Fig 4. Graph illustrating the results of the second experiment.

2) Results & Discussion

An analysis of the combined fifth and sixth sessions of the online shopper shows a much-improved predictive accuracy for the online customer with the same threshold value (in or around .75) and a performance that is marginally better than that of the real-world customer. This effect is not the result of any double counting of predicted items but rather is based on the proportion of unique items in the combined lists that were accurately predicted. So using this more complete online list, based on two weeks of shopping, we see that the predictive accuracy can be at least as good as that of the real-world case. Having said this, it is still probably best to rely on real-world sessions in building user profiles, as they will tend to be more accurate in a given week.

V. CONCLUSIONS AND FUTURE WORK

We have presented a system that attempts to reinstate the external memory of a store for an online shopper, making use of the customer's purchasing history and the physical layout of the actual store at which they shop. The system is based on a novel idea we called *Memory Zones*. We have seen that the system can be calibrated to make reasonable recommendations to overcome an online shopper's forgetting of items to be purchased. This system will be most useful in application domains where people continue to shop, partially online and partially in the actual store. We also see extensions to this system that will enable more targeted recommendation to online customers.

There are, of course, other extensions that could be made. For example, at present, the system has no notion of the quantity of purchases. For example, a user buys one jar of coffee and does not buy coffee again for two months. If the next time that that user buys coffee he/she buys two

jars, then it stands to reason that those two jars should last twice as long as one jar. There is no concept of quantity of groceries in the system. Another possible 'would be to automatically generate the threshold value for adding items to the memory zone. This threshold is currently static and a dynamic generation could take into account for example instances where it has been a number of weeks since the customer last visited the store for example. If these ideas were implemented and taken into account then recommendations could potentially become even more accurate.

REFERENCES

- [1] http://www.ananova.com/business/story/sin_672180.html?menu=
- [2] <http://www.checkout.ie/WorldReport.asp?ID=5>
- [3] <http://news.bbc.co.uk/2/hi/business/2571695.stm>
- [4] Eysenck MW, Keane MT. Cognitive Psychology : A Students Handbook (4th Edition) London Psychology Press pp 630, 2000.
- [5] Ahlberg C, Williamson C, Schneiderman B : Dynamic Queries for information Exploration : An Implementation and Evaluation, Proceedings of ACM CHI, Human Factors in Computer Systems, 1992.
- [6] Domingue J, Martins M, Tan J, Stutt A, Pertusson H : ALICE : Assisting Online Shoppers through Ontologies and Novel Interface Metaphors. 13th International Conference on Knowledge Engineering and Knowledge Management, Sigüenza (Spain), October 2002.
- [7] Cotter P, Smyth B : PTV: Intelligent Personalised TV Guides, Proceedings of the 12th Innovative Applications of Artificial Intelligence Conference, AAAI Press, 2000.
- [8] Lawrence RD, Almasi GS, Kotlyar V, Viveros MS, Duri SS. Personalisation of Supermarket Product Recommendations, Data Mining and Knowledge Discovery 5, pp 11-32, 2001.
- [9] David Goldberg, David Nichols, Brian M. Oki, Douglas Terry. Using Collaborative Filtering to Weave an Information Tapestry. Communications of ACM, vol. 35, no. 12, pp. 61-67, 1992
- [10] Paul Resnick, H.R. Varian. Recommendation Systems. Communications of ACM, vol. 40, no. 3, pp. 56-58, 1997.
- [11] www.amazon.com
- [12] R Agrawal, T. Imielinski, A. Swami: Mining Associations between Sets of Items in Massive Databases, Proc of the ACM-SIGMOD International Conference on Management of Data, Washington D.C., pp. 207-216 May 1993.
- [13] R. Agrawal, R. Srikant: Fast Algorithms for Mining Association Rules, Proc. Of the 20th International Conference on Very Large Databases, Santiago, Chile, Sept. 1994.
- [14] Maurice Houstma, Arun Swami. Set-Oriented Mining of Association Rules. Proceedings of International Conference on Data Engineering, 1995
- [15] Shardanand, Upendra and Maes, Pattie. Social Information Filtering : Algorithms for Automating "Word of Mouth". Proceedings of Computer Human Interaction, pp. 210-217, 1995.
- [16] Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom, John Riedl. GroupLens: An Open Architecture for Collaborative Filtering of Netnews. Proceedings of ACM Conference on Computer Supported Cooperative Work, pp. 175-186, 1994.
- [17] Smyth B, Cotter P. The Skys the Limit : A Personalised TV Listings Service for the Digital Age, Proceedings of the 19th SGAI International Conference on KBS and Applied AI (ES) Cambridge UK, 1999.
- [18] Raymond J. Mooney, Loriene Roy. Content Based Book Recommending Using Learning for Text Categorization. Proceedings of D1-00, 5th ACM Conference on Digital Libraries.
- [19] Levy M, Weitz B. Retail Management (3rd Edition) Irwin/McGraw - Hill, 1998.