Analysis of Recommender System Usage by Multidimensional Scaling

Patrick Thoma and Wolfgang Gaul

Institut für Entscheidungstheorie und Unternehmensforschung,
Universität Karlsruhe (TH), 76128 Karlsruhe, Germany

Abstract. Recommender systems offer valuable information not only for web site visitors (who are supported during site navigation and/or buying process) but also for online shop owners (who can learn from the behavior of their web site visitors). We use data from large German online stores gathered between March and November 2003 to visualize search queries by customers together with products viewed most frequently or purchased most frequently. Comparisons of these visualizations lead to a better understanding of searching, viewing, and buying behavior of online shoppers and give important hints how to further improve the generation of recommendations.

1 Introduction

A recommender system can be defined as software that collects and aggregates information about site visitors (e.g., buying histories, products of interest, hints concerning desired/desirable search dimensions or FAQ) and their actual navigational and buying behavior and returns recommendations (e.g., based on customer demographics and/or past behavior of the actual visitor and/or user patterns of top sellers with fields of interest similar to those of the actual contact (Gaul and Schmidt-Thieme (2002)). A framework to classify such systems according to their input and output facilities can be found in Gaul et al. (2002). A substantial problem in the evaluation of recommender systems is that in real-life environments their influence to customers buying behavior can hardly be measured isolated from other effects, e.g., promotions, price reductions, etc. The research questions therefore was to analyze the acceptance and the functioning of recommender systems in a more qualitative way by making the site visitors interaction with it transparent.

In the following we will use multidimensional scaling (MDS) to visualize how online shoppers place search queries and react to recommender system output where MDS is the label for a class of methods that represent similarity or dissimilarity values with respect to pairs of objects as distances between corresponding points in a low-dimensional metric space (Borg and Groenen (1997)). The graphical display of the object representations as provided by MDS enables to literally "look" at the data and to explore structures visually. We used this technique to analyze recommender system usage on the basis of two data sets, collected from two different German online retail stores.

Both use a ruled-based recommender system with the same kernel functionality to support customers during the buying process. As a result relations between searching, viewing and buying behavior of the corresponding site visitors visualized in the underlying product space revealed valuable insights for product managers and recommender system engineers.

2 Methodology

A popular and classical technique to construct object representations in a low-dimensional space is MDS (Kruskal (1964)). The goodness of fit of the solution obtained within the different iterative steps of whatever method can be assessed by the so called *Kruskal stress*. We used an implementation of Kruskal's non-metric MDS which is available as the *isoMDS* function in the "MASS" library (Venables and Ripley (1997)) of the "R" software package. To overcome a weakness of classical MDS, i.e. the interpretation of the object positionings in low-dimensional spaces, we applied *property fitting*.

Let $O = \{o_1, \dots, o_N\}$ be the set of objects and $b_n = (b_{n1}, \dots, b_{nM})$ the representation of object o_n in the underlying M-dimensional target space, $n = 1, \dots, N$. If additional information about the objects is given, e.g., in form of attribute vectors $a_p = (a_{1p}, \dots, a_{Np})'$ for property p, $p = 1, \dots, P$, one can construct property vectors $c_p = (c_{p1}, \dots, c_{pM})$ in the M-dimensional space so that the projections

$$\hat{a}_{np} = \sum_{m=1}^{M} c_{pm} b_{nm} \tag{1}$$

of b_n onto c_p approximate the actual attribute values of the objects as good as possible with respect to the least squares criterion

$$\sum_{n=1}^{N} (\hat{a}_{np} - a_{np})^2. \tag{2}$$

Vector notation leads to $c_p = (B'B)^{-1}B'a_p$ with $B = (b_{nm})$. Quality of fit can be measured by correlation coefficients between $\hat{a}_{np}$ and a_{np}.

Similarly, we performed the subsequent transformation of the search queries $s_q = (s_{q1}, \dots, s_{q\tilde{P}_q})'$. Here, $\tilde{p} = 1, \dots, \tilde{P}_q, \tilde{P}_q \leq P$, indicates properties specified in the underlying query. In cases where a range of values was stated for a property p, e.g., for the price, we set $s_{q\tilde{p}}$ equal to the mean obtained from the lower and upper boundaries of the specified range. With given property vectors $c_{\tilde{p}}$ we looked for the representation $z_q = (z_{q1}, \dots, z_{qM})'$ of s_q so that the projections

$$\hat{s}_{q\tilde{p}} = \sum_{m=1}^{M} z_{qm} c_{\tilde{p}m} \tag{3}$$

of z_q onto $c_{\tilde{p}}$ approximate the actual attribute values of the search profiles as good as possible with respect to the least squares criterion

$$\sum_{\tilde{p}=1}^{\tilde{P}_q} (\hat{s}_{q\tilde{p}} - s_{q\tilde{p}})^2. \tag{4}$$

Vector notation leads to $z_q = (C_q' C_q)^{-1} C_q' s_q$ with $C_q = (c_{\tilde{p}m})$.

3 Empirical results

3.1 The data sets

We collected system usage information from two large German online retail stores. The first data set contains searching, viewing and buying information of website visitors looking for notebooks, the second data set contains the respective information for washing machines. Both stores support their customers with the help of recommender systems for dedicated product domains. In both cases, users can specify a search query by defining the importance of and the desired value for every attribute from a list of given attributes. The recommender systems - as response - compute sorted lists of products which best fulfill the customers requirements using an internal ruled-based algorithm. The quality of the proposals is implicitly evaluated by counting the number of clicks on the images of the suggested products which lead to pages with additional, more detailed information about the products of interest and additionally, by counting how often corresponding products were put into electronic market baskets. These two events are in the following called "views" and "purchases".

The data set about notebooks was gathered between March and July 2003. Products were described by a list of 14 attributes of which we selected price, clockrate, ram, harddisk, display, drives, weight, interfaces, battery, and software for our analysis. Using these 10 properties a list of 307 different products could be recommended. The data set contained 7125 search queries of which 434 were empty which means that no values were specified for any attribute. Feedback information consisted of 15305 views and 509 purchases.

The data set for washing machines was collected from May until November 2003. In this case, the selected attributes were price, type (front or top loader), charge, effectiveness, maximum spin, programs, and extras. This data set contains 54 different products, 20024 search queries (thereof 14131 empty ones), 49696 views, and 758 purchases.

A comparison of these numbers shows the following: Online shoppers had a closer look at 2-3 products (notebooks: 2.15, washing machines: 2.50) per search query. The different number of empty search queries and the fact, that for notebooks, the "click-to-buy-ratio" (purchases/views) is more than twice the number for washing machines (notebooks: 3.3%, washing machines: 1.5%)

might be an indicator that site visitors looking for notebooks have already a clearer conception of what they are searching and thus are more willing to buy online than users looking for washing machines. In our analysis we tried to verify these assumptions of this kind.

3.2 Representation of products and search profiles

A first step of our analysis was to display the products together with the specified search queries in a two-dimensional space. Dissimilarities between objects were defined by first scaling all attribute vectors to $N(0,1)$ and then calculating euclidean distances using the *dist* function (The R development core team (2003)). As mentioned above, the object representations were obtained with the help of the *isoMDS* function. The *stress* measures of the so found solutions were sufficient (0.175) for notebooks and good (0.146) for washing machines. The correlation coefficients of the property vectors where rather high (> 0.7) in most cases, exceptions are interfaces (0.47) and battery (0.56) for notebooks and programs (0.56) for washing machines. To structure the many search queries, we grouped identical queries together and used the notation "search profiles" for identical ones. We selected search profiles with at least 10 queries and transformed them into the two-dimensional target space according to the methodology explained in section 2. The results are shown in figures 2 and 1. In figure 2 one can see, that notebooks and search profiles are relatively equally spread around the origin. From a management's point of view, this could be interpreted as hint that the product catalog for notebooks covers the range of customers' requirements quite well. This is not always the case as Gaul et al. (2004) revealed when analyzing a recommender system for digital cameras. Remarkable is always the group of search profiles forming a straight configuration in direction of the negative prolongation of the price vector (The price vector is marked with a dashed line.). These profiles stand for queries where a low price was the only specified attribute. From the comparatively high number of such profiles one can – not surprising – conclude that price is one of the most important search criteria and that many of the users are quite price sensitive since these profiles represent demand for products below the average price. For washing machines, figure 1 shows two more distinguishable lines of search profiles. These configurations result from queries that are identical in all but one attribute. The differentiating attribute is the number of extras like a "water-stop-function" or a "timer". Contrary to the notebooks' data set, most of the search profiles are situated in the area of above average prices. This could be a hint that online shoppers interested in washing machines are less price sensitive and more in favor of additional equipment of the products.

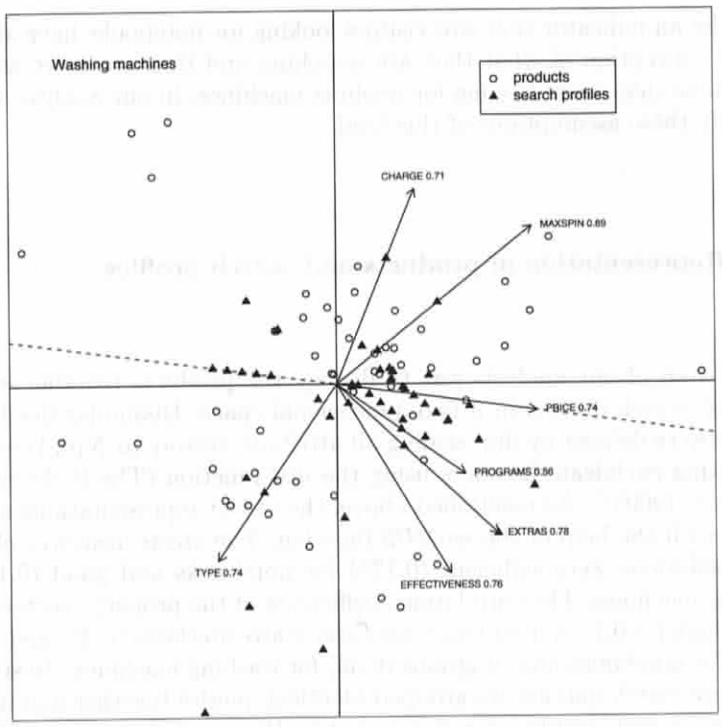

Fig. 1. Property vectors, search profiles, and products (Washing machines)

3.3 Analysis of system usage

In the second step, we analysed coherences between specified search profiles and subsequently viewed and/or purchased products. For this purpose, we determined for each profile the ten most frequently viewed products and connected their object representations in the target space with the representation of the search profile. This would lead to many spider like graphs. In figure 3 the situation is shown for two selected search profiles from the data set for washing machines. In both cases, most of the frequently viewed products are relatively close to the specified search profile. However, there are also products which have been viewed frequently but are positioned quite far away from the originally specified search profile. For the search profile in the 3rd quadrant these are the two products in the 4th quadrant. Looking at the underlying data shows that these two products are of different type (front- instead of top-loader). The opposite case also exists: There are products which seem to fulfill the requirements quite well according to their positioning in the spider graph but have not been viewed frequently. As users can only view products that have been recommended, this could be an indicator that these products were not on the recommendation list for the given search profile. Checking

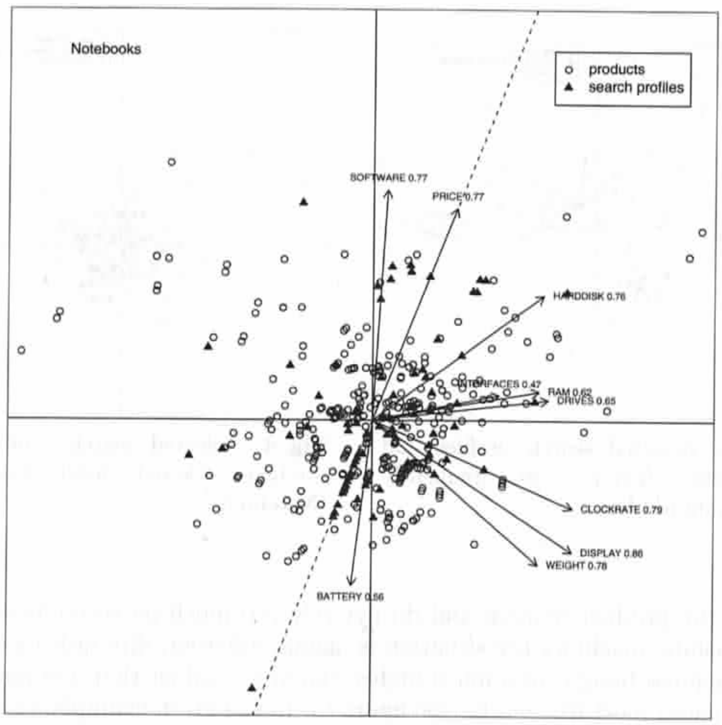

Fig. 2. Property vectors, search profiles, and products (Notebooks)

this with the actual recommendations - an information not available in our data sets - might help to improve recommendation generation.

The same analysis for notebooks showed a different picture (figure 4). In this case, the most frequently viewed products cannot really be characterized as always close to the search profile. From the underlying data, we can see that a desired price of 900 EURO was specified while the prices of the products viewed most frequently ranged from 979 EURO up to 1649 EURO. This means that online shoppers looking for notebooks are restrictive with respect to price when formulating their search queries, but are nevertheless interested in better equipped items - even if the price is far beyond the their preferred value.

This leads to the question how site visitors decide when it comes to buying. Figure 5 shows that for the selected search profile there is only one product in the category "purchased most frequently". Remarkably enough, this product is not among the ones of the category "viewed most frequently". Its representation has a medium distance from the representation of the search profile (in numbers: its price was 1249 EURO while 900 EURO were specified in the search query). This supports the assumption from section 3.1 that online shoppers looking for notebooks have already quite a good knowledge

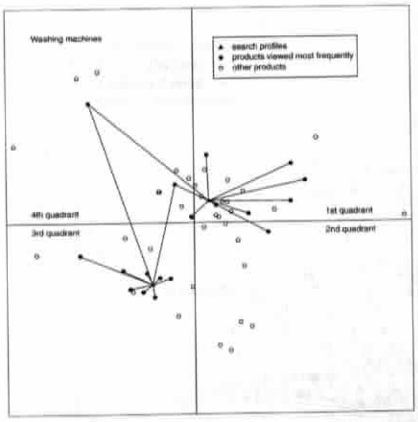

Fig. 3. Selected search profiles and products viewed most frequently (Washing machines)

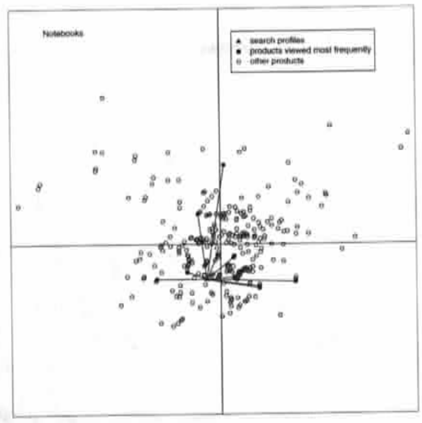

Fig. 4. Selected search profiles and products viewed most frequently (Notebooks)

about the product domain and do not rely too much on recommendations. For washing machines the situation is, again, different. Site visitors of washing machines bought to a much higher extent a product that was among the ones viewed most frequently (see figure 6). In the given example, no product was purchased after the search query in the 3rd quadrant had been specified. The products purchased on basis of the search query in the 1st quadrant are all relatively close and had been viewed frequently. This could be seen as a confirmation, that online shoppers looking for washing machines are more uncertain about their actual needs and therefore more willing to rely on suggestions of recommender systems.

4 Summary

In this paper two product domains were used to demonstrate, that Multidimensional Scaling can be a suitable tool to analyze recommender system usage. For the given data sets, products were represented in a two-dimensional space with sufficient/good stress measures and high correlation coefficients for additional property vectors. The presented methodology allowed us to subsequently transform the search profiles into the same target space and to display them together with the products in joint representations. The visualization of searching, viewing and buying behaviour of online shoppers showed remarkable differences between the two product domains. While notebook site visitors seem to have extensive knowledge concerning the product domain and to be rather independent from system recommendations, online shoppers looking for washing machines are more uncertain and therefore more willing to rely on recommender system's suggestions when they make their

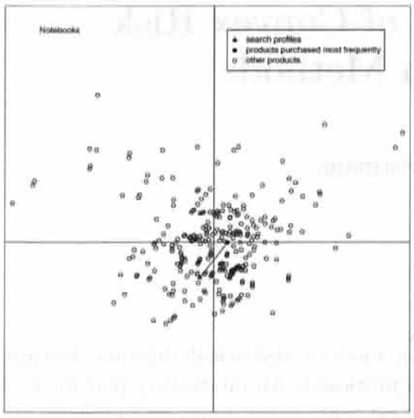

Fig. 5. Selected profiles and products purchased most frequently (Notebooks)

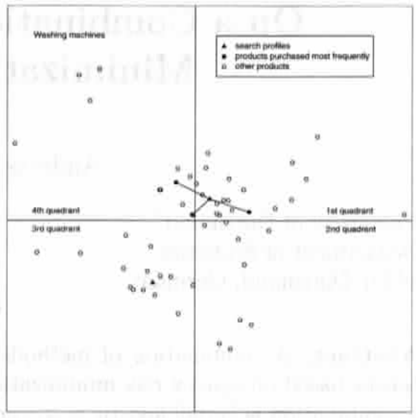

Fig. 6. Selected search profiles and products purchased most frequently (Washing machines)

buying decision. Analysis of this kind can create value for product managers and shop owners as well as for developers of recommender systems' software.

References

BORG, I. and GROENEN, P. (1997): *Modern Multidimensional Scaling: Theory and Applications.* Springer, New York.

GAUL, W., GEYER-SCHULZ, A., SCHMIDT-THIEME, L. and HAHSLER, M. (2002): eMarketing mittels Recommendersystemen. *Marketing ZFP, 24, 47–55.*

GAUL, W. and SCHMIDT-THIEME, L. (2002): Recommender Systems Based on User Navigational Behavior in the Internet. *Behaviormetrika. 29, 1–22.*

GAUL, W., THOMA, P., SCHMIDT-THIEME, L., and VAN DEN BERGH, S. (2004): Visualizing Recommender System Results via Multidimensional Scaling, to appear in: *Proceedings of OR 2003 International Conference.* Springer, New York.

KRUSKAL, J. B. (1964): Multidimensional Scaling by Optimizing Goodness of Fit to a Nonmetric Hypothesis. *Psychometrika, 29, 1–27.*

THE R DEVELOPMENT CORE TEAM (2003): *The R Environment for Statistical Computing and Graphics.*
http://cran.r-project.org/doc/manuals/fullrefman.pdf

VENABLES, W.N. and RIPLEY, B.D. (1997): *Modern Applied Statistics with S-PLUS*, 3rd ed. Springer, New York.