Mining Retail Data for Targeting Customers with Headroom

Madhu Shashanka, Michael Giering
Mars Inc., 100 International Drive, Mount Olive NJ 07828

Recommender systems have recently gained a lot of attention both in industry and academia. In this paper, we focus on the applications and utility of recommender systems for brick-and-mortar retailers. We present a case study of a project that we completed for a large retail chain whose goal was to mine the transaction data to understand shopping behavior and target customers who exhibit *headroom* - the unmet spending potential of a shopper in a given retailer. There are two aspects to the problem - identifying the shoppers with headroom and product categories where they can spend more, and estimating shoppers' unrealized spending potential in product categories that they haven't bought.

Data from every transaction from over 350 stores of a large retail chain gathered over a period of 16 months was provided to us. Data was restricted to transactions of regular shoppers who used a "loyalty card" that could track them across multiple purchases.

We employed the technique of Singular Value Decomposition (SVD) for analysis. SVD is fundamental to many data modeling and data mining algorithms. SVD factorizes an $M \times N$ matrix $\mathbf{X}$ into two orthogonal matrices $\mathbf{U}$, $\mathbf{V}$ and a diagonal matrix $\mathbf{S} = diag(\mathbf{s})$ such that $\mathbf{USV}^T = \mathbf{X}$ and $\mathbf{U}^T\mathbf{XV} = \mathbf{S}$. A *thin*-SVD $\mathbf{U}'\mathbf{S}'\mathbf{V}'^T$ is the optimal rank-k approximation of $\mathbf{X}$ in the least-squares sense where only the k largest singular values and corresponding singular vectors are retained in $\mathbf{S}'$, $\mathbf{U}'$, $\mathbf{V}'$.

Several well-known recommender systems are based on SVD (e.g. [1, 2]). Let the matrix $\mathbf{X}$ represent a matrix of consumer spends over a given period, where columns represent different shoppers and rows represent different product categories. The subspace spanned by the columns of $\mathbf{U}'$ in a thin-SVD can be interpreted as the k most important types of "shopping profiles" and a location can be computed for each shopper in this *shopping profile space*. The relationship between $\mathbf{x}_n$, the n-th column of $\mathbf{X}$ representing the spends of the n-th shopper, and his/her location $\mathbf{p}_n$ in the shopping profile space is given by $\mathbf{p}_n = \mathbf{U}'^T\mathbf{x}_n$. This vector $\mathbf{p}_n$, equal to the n-th row of the matrix $\mathbf{V}'\mathbf{S}'$, underlies all SVD-based recommender systems. The idea is to estimate $\mathbf{p}_n$ and thus obtain imputed values for missing data in $\mathbf{x}_n$. Because of the large size of the data set and unsuitability of traditional SVD algorithms for large data sets, we used the iterative incremental SVD implementation (IISVD) of Brand [1] which can handle large data sets with missing values.

All variables in the data-set such as total spends, total number of shopping trips, and number of items bought showed strong log-normality, making SVD an ideal tool for analysis. Data preprocessing steps included elimination of outliers, log-normalization and centering.

Shopping behavior profiles were generated by expressing each shopper's cumulative spend for each product category in percentage terms (considering percentages masks the effect of shopping household size on the magnitude of spend and focuses on the relative spend across categories) using a thin-SVD $\mathbf{U}'\mathbf{S}'\mathbf{V}'^T$. Our experience shows that one can reasonably expect a noise content of about 15% in retail data. Using this as a rough threshold, we computed the rank as 26 by keeping only as many singular vectors whose cumulative variance measures sum to less than 85% of the total data variance. We then segmented the shoppers by running K-means clustering on the *shopper locations* given by the 26-dimensional rows of matrix $\mathbf{V}'\mathbf{S}'$, resulting in 23 distinct clusters of shoppers.

Within a given cluster of shoppers, let $\mathbf{X}$ represent the matrix of spends where the two dimensions represent shoppers and product categories. We used IISVD for computing full-rank decompositions and obtained imputed values for all the missing data in each of the 23 clusters. These imputed spends for a given shopper are equivalent to a linear mixture of the spends of all shoppers within the cluster, weighted by the correlations between their spends and spends of the current shopper.

For the identification of shoppers with headroom and corresponding product categories, we needed to measure how each shopper is over-spending or under-spending in each category. Using the filled shopper spend data from the previous step, we removed 10% of the known data values and re-imputed them using a thin-SVD algorithm. The thin-SVD algorithm determined the optimal rank for minimizing the error of imputation by way of cross-validation. This process was repeated to re-impute all the known values. The differences between the actual and modeled spends of known values, being normally distributed, can be expressed as Standard Z-scores. These Z-scores can be interpreted as a probabilistic measure of shopper over-spending/under-spending across product categories, thus indicating headroom. Also, we expressed actual spend data and *spend-per-item* data in terms of Z-scores for each category, which could be used as proxies related to headroom. These metrics, along with shopping frequency, provide retailers with the ability to identify product categories most relevant to shoppers and persuade them to spend to their potential.

Due to the customers' demand for mass customization in recent years, it has become increasingly critical for retailers to be a step ahead by better understanding consumer needs and by being able to offer promotions/products that would interest them. The system presented in this paper is a first step in that endeavor. We believe that with a few iterations of this process and fine tuning based on feedback from sales and promotions performance, it can be developed into a sophisticated and valuable retail tool.

[1] M Brand. Fast Online SVD Revisions for Lightweight Recommender Systems. In *SIAM Intl Conf on Data Mining* 2003.

[2] B Sarwar, G Karypis, J Constan, J Reidl. Application of Dimensionality Reduction in Recommender System - A Case Study. In *ACM WebKDD*, 2000.