Simultaneous Co-clustering and Modeling of Market Data

Meghana Deodhar and Joydeep Ghosh

Department of Electrical and Computer Engg.
University of Texas at Austin, Austin, TX, USA.

Abstract. For difficult prediction problems, practitioners often segment the data into relatively homogenous groups and then build a model for each group. This two-step procedure usually results in simpler, more interpretable and actionable models without any loss in accuracy. We consider two important marketing problems, predicting customer-product preference and simultaneous market segmentation and structure. We present a model-based co-clustering algorithm that interleaves clustering of customers and products and construction of prediction models to iteratively improve both cluster assignment and fit of the models. Our approach applies to a wide range of bi-modal or multimodal market data, where it can be used to address prediction or segmentation problems. We demonstrate the effectiveness of our approach on the ERIM market data.

1 Introduction

For many difficult prediction problems, a single "global" model is not very suitable. Instead, it is advantageous to partition the population into multiple, relatively homogeneous segments and then develop separate models for each segment [1, 2]. For example, while forecasting electric load usage, it is advisable to build separate predictive models for weekdays, weekends and holidays. This divide-and-conquer approach not only improves the overall prediction accuracy but also interpretability, since the component models are often far simpler [3].

In this paper we consider the problem of predicting customer-product preferences. The dataset in this case is a matrix of customers by products, with cells representing the corresponding customer-product preference values, which could be ratings, choice probabilities or the number of units of the product purchased by the customer. This matrix will have missing values where the corresponding customer-product preference is unknown. Each customer is described by a set of attributes, for example, demographics and each product also has attributes, which could include the price, market share, quality, etc. The problem is to predict the preference for the missing customer-product combinations, as well as behavior of new customers, or preferences for new products.

Collaborative filtering approaches for this problem will make use only of the matrix entries and ignore customer/product attributes [4, 5]. On the other extreme, a typical prediction model will form a map between the independent or explanatory variables (customer and product attributes) for a given customer-product pair and the dependent variable (the corresponding matrix entry), but will not consider "nearby" customers or products in this process. For a diverse population of customers and a wide range of different products, it is unlikely that all the customer-product preferences can be well explained by a single model. It is more natural for a model to closely represent the preferences of only a subset of the customers for a subset of the products.

Another technique that can be applied to this problem is co-clustering, which simultaneously clusters the data matrix along multiple axes and has been successfully applied in several domains like text clustering and microarray data analysis [6]. Co-clustering exploits the duality between the two axes to improve on traditional "single-sided" clustering (i.e. segmentation). A co-clustering approach will simultaneously cluster the customers and products based on the matrix entries. It will then use the entries of the corresponding co-cluster to predict a missing value [5]. However, like collaborative filtering, this approach also ignores the customer and product attributes, i.e., the prediction is solely based on the value of the dependent variable in a suitably identified neighborhood.

We propose an approach that exploits *both* neighborhood information as well as the available customer/product attributes. The idea is to co-cluster the entire data matrix into blocks of customers and products such that each block can be well characterized by a single predictive model. Our *model based co-clustering-cum-learning* algorithm achieves this by interleaving clustering and construction of prediction models to iteratively improve both cluster assignment and fit of the models. This simultaneous approach is better than independently clustering the data first and then building predictive models. We address the specific application of predicting the number of items of a given product purchased by a customer, by building regression models in each co-cluster. We apply this technique to a formidable, real marketing dataset and show the effectiveness of our approach in Section 6.

Another application of our model based co-clustering framework is simultaneous market segmentation and market structure [7]. Market segmentation involves dividing the market into groups of consumers who are homogenous in terms of their choice of products. Market structure deals with clustering products into groups such that the products within each group are equivalent in some respect, and can be used to find groups of competing products in a given market. Both these problems are typically addressed separately by most researchers, without exploiting the duality between them. The model based co-clustering framework is promising for this problem since it simultaneously clusters customers and products and builds explanatory models for each co-cluster, which indicate the factors that influence the preference for a group of products. We address this application in Section 6.2.

Notation: Small letters represent scalars e.g. a,z, bold face letters represent vectors e.g. $\mathbf{b}, \mathbf{c}, \boldsymbol{\beta}$, capital letters like Z, W represent matrices. Individual elements of a matrix e.g. Z are represented as z_{ij}, where i and j are the row and column indices respectively.

2 Related Work

Wedel and Steenkamp proposed a generalized fuzzy clusterwise regression technique to simultaneously perform market segmentation and structure [8]. Each cluster includes fractional membership from all customers and products and is hence a fuzzy co-cluster. Each cluster has a regression model that predicts the preferences as a linear combination of the product attributes. The cluster memberships and models are estimated so as to reduce the total squared error between the actual preferences and the predicted preferences. However, the hard version of this method corresponds to diagonal co-clustering [6], where only a small subset of products are associated with each customer group. Such a partitioning does not give information about all possible customer-product combinations. In contrast, this paper is concerned with partitional co-clustering, which covers the entire customer-product matrix.

Other approaches to simultaneous market segmentation and structure include the work by Grover and Srinivasan, who use a cross classification matrix of the proportion of customers that switch from one brand to another [7]. Maximum likelihood is used to estimate latent class models that can be used to extract consumer segments. Competing products for each segment are determined by identifying the brands with large purchase probabilities in that segment. Moe and Fader model customer and product segments simultaneously in their analysis of music CD sales [9]. The sales for each product segment are modeled as an additive combination of contributions from different consumer segments, each of which have different purchase parameters.

3 Problem Definition

The data is represented as an $m \times n$ matrix Z of m customers and n products, with cells z_{ij} representing the corresponding customer-product preference e.g. the number of units of product j purchased by customer i. A weight w_{ij} is associated with each cell z_{ij}, ordinarily set to 1. Missing values are dealt with by setting their weights to 0. Less certain data values can be handled by giving them comparatively lower but non-negative weights.

A customer i has attributes $\mathbf{C_i}$, and product j has attributes $\mathbf{P_j}$. It is assumed that each matrix value z_{ij} is generated by a linear model involving these attributes, where the preference value $z_{ij} \in \mathbf{R}$ is modeled as a linear combination of the corresponding customer and product attributes. The preference value is estimated as $\hat{z}_{ij} = \beta_0 + \boldsymbol{\beta_c}^T \mathbf{C_i} + \boldsymbol{\beta_p}^T \mathbf{P_j}$. The aim is to simultaneously cluster the customers and products into a grid of k row clusters and l column clusters, such that preference values within each co-cluster have similar linear models and can be represented by a single common model. The co-cluster assignments along with the regression models for the co-clusters can also be used to predict unknown customer product preference values.

Formally, let ρ be a mapping from the m rows to the k row clusters and γ be a mapping from the n columns to the l column clusters. We want to find a co-clustering defined by (ρ, γ) and the associated set of $k \times l$ regression models

$\{\boldsymbol{\beta}^{\boldsymbol{gh}}\}$ that minimize the following objective function

$$\sum_{g=1}^{k}\sum_{h=1}^{l}\sum_{u:\rho(u)=g}\sum_{v:\gamma(v)=h} w_{uv}(z_{uv}-\hat{z}_{uv})^2, \quad (1)$$

where z_{uv} is the original value in row u, column v of the matrix, with associated weight w_{uv} and $\hat{z_{uv}} = \boldsymbol{\beta}^{\boldsymbol{gh}^T}\mathbf{x_{uv}}$ is the predicted value. Here $\boldsymbol{\beta}^{\boldsymbol{gh}} = [\beta_0^{gh}, \boldsymbol{\beta}_{\boldsymbol{c}}^{\boldsymbol{gh}^T}, \boldsymbol{\beta}_{\boldsymbol{p}}^{\boldsymbol{gh}^T}]^T$ denotes the vector of coefficients of the model associated with the co-cluster that the cell value z_{uv} is assigned to and $\mathbf{x_{uv}} = [1, \mathbf{C_u}^T, \mathbf{P_v}^T]^T$ is a vector consisting of customer and product attributes. Since the weights for the missing z_{uv} values are 0, the cost function essentially ignores them and is simply the squared error summed only over all the known elements of matrix Z.

4 Simultaneous Co-clustering and Regression

Since the objective function (1) is the squared error summed over all the elements of the matrix, it can be expressed as a sum of row or column errors. If row u is assigned to row cluster g (i.e. $\rho(u) = g$), the row error is $E_u(g) = \sum_{h=1}^{l}\sum_{v:\gamma(v)=h} w_{uv}(z_{uv} - \boldsymbol{\beta}^{\boldsymbol{gh}^T}\mathbf{x_{uv}})^2$. For a given column clustering and model parameter sets $\{\boldsymbol{\beta}^{\boldsymbol{gh}}\}$, the best choice of the row cluster assignment for row u is the g that minimizes this error, i.e., $\rho^{new}(u) = argmin_g E_u(g)$. A similar approach is used to (re)-assign columns to column clusters. Such row and column cluster updates hence decrease the objective function and improve the clustering solution.

Given the current row and column cluster assignments, the co-cluster models need to be updated, i.e. the coefficient vector $\boldsymbol{\beta}$ has to be updated for each co-cluster. The model for a row cluster g of size r and column cluster h of size c is updated by solving $\min\|\mathbf{w}_{(r*c)\times 1}\mathbf{z}_{(r*c)\times 1} - X_{(r*c)\times p}\boldsymbol{\beta}_{p\times 1}\|_2^2$, where $\mathbf{z}$ is a vector of all the $r*c$ preference values in the co-cluster and $\mathbf{w}$ is a vector of the corresponding weights. X is the matrix of the corresponding row and column attributes. p is the total number of coefficients to be estimated, including the intercept. The $\boldsymbol{\beta}$ is hence a solution to a weighted least squares problem, that minimizes the sum of the squared errors between the original values and the predicted values. The model update step is hence guaranteed to decrease the objective function.

The resulting algorithm is a simple iterative algorithm described in Figure 1. Step 1 minimizes the objective function due to the property of linear regression, steps 2(a) and 2(b) directly minimize the objective function. The objective function hence decreases at every iteration. Since this function is bounded from below by zero, the algorithm is guaranteed to converge to a local minimum.

Predicting missing preference values. Let z_{uv} be a missing cell value that has been assigned to row cluster g and column cluster h. $\mathbf{x_{uv}}$ is the vector of attributes of row u and column v and $\boldsymbol{\beta}^{\boldsymbol{gh}}$ represents the model parameters of the linear regression model of the assigned co-cluster. The missing value z_{uv} is predicted as $\hat{z_{uv}} = \boldsymbol{\beta}^{\boldsymbol{gh}^T}\mathbf{x_{uv}}$.

```
Algorithm
Input: Z_{m×n}, W_{m×n}, C = [C_1..C_m] , P = [P_1..P_n]
Output: Co-clustering (ρ, γ) and co-cluster models β's
1. Begin with a random co-clustering (ρ, γ)
2. Repeat
3.        Step 1: Update co-cluster models
4.        for g = 1 to k do
5.            for h = 1 to l do
6.                Use linear regression to update the coefficient vector β_gh
7.            end for
8.        end for
9.        Step 2(a): Update ρ
10.       Assign each row to the row cluster that minimizes the row error
11.       Step 2(b): Update γ
12.       Assign each column to the column cluster that minimizes the column error
Until Convergence
13. return (ρ, γ) and β's
```

Fig. 1. Pseudo-code for simultaneous co-clustering and regression

5 Reduced Parameter Approach

The simultaneous co-clustering and regression approach requires a total of $(1 + |C| + |P|) \times kl$ parameters, where $|C|$ and $|P|$ are the number of customer and product attributes respectively, since it constructs $k \times l$ independent models, one per co-cluster. This model may overfit in cases where training data is limited due to the presence of too many parameters. The other extreme is a single prediction model for all the data, with $(1 + |C| + |P|)$ parameters, which might not be adequate. We propose an intermediate approach, which constructs $k \times l$ models but with regularization achieved by sharing parameters across certain sets of models. The customer coefficients of all models in the same row cluster and the product coefficients of all models in the same column cluster are constrained to be identical. This reduces the number of parameters to $(1+|C|) \times k + (1+|P|) \times l$. For more details of the algorithm for the reduced parameter approach, the reader is referred to [10].

6 Experimental Results on a Real Marketing Dataset

We applied the simultaneous co-clustering and regression approach to two challenging marketing applications,(i) predicting unknown customer-product preferences and (ii) simultaneous market segmentation and structure. The dataset that we use is the publicly available ERIM dataset [1] consisting of household panel data collected by A.C. Nielsen, which has been used by several researchers [11–13]. This dataset has purchase information for six product categories over a

[1] URL: `http://www.gsb.uchicago.edu/kilts/research/db/erim/`

period of 3 years from 1985-1988 for households in Sioux Falls, South Dakota. The dataset includes household demographics as well as product characteristics.

The data preprocessing steps we took are similar to the data selection procedure by Seetharam et al. [13]. We have 6 product categories (ketchup, tuna, sugar, tissue, margarine and peanut butter) with a total of 121 products (brands). Brands with very low market share in each product category are omitted. We select households that made at least 2 purchases in each product category, resulting in a set of 1714 households. We select 6 household attributes - income, number of members, male head employed, female head employed, # shopping visits and total amount spent and 3 product attributes - market share, price, # times advertised. The data can be represented by a data matrix of households and brands where the cell values are the number of units of a brand purchased by a household, aggregated over the time the household was tracked. The number of units purchased can be used as an indicator of brand preference.

Dataset Properties. The data matrix is extremely sparse, with 74.86% of the values being 0. The distribution of the number of units purchased is very skewed. 99.12% of the values are below 20, while the remaining values are very large and range upto around 200. These few, large values that form the tail in the histogram of the matrix entries, can be considered as outliers with respect to the rest of the values. Linear "least squares" regression is sensitive to outliers, and a few outliers could skew the model results. Hence on this dataset, for a fair comparison of linear model based techniques with respect to prediction of missing values, we need some way of dealing with outliers. It is unreasonable for a model based on linear regression to capture both small as well as extremely large values simultaneously and a more suitable approach would be to separate out these two very different sets of values and model them independently. A threshold of 20 for the number of units purchased was used to separate the bulk of the matrix entries (99.12%) from the tail of high values. In this paper we focus on the model for the bulk of the entries, the results of which are illustrated in Section 6.1.

Standardization of the Data. The products in this application are from 6 different product categories, with price and extent of advertising varying with category. When we construct a linear model for a co-cluster we weigh the attributes of all the products in the co-cluster by the same set of coefficients. However, the products in the co-cluster could be from different categories with very different ranges of attribute values. We hence need to standardize the product attributes to make them comparable across categories. We transform each product attribute value a to $a^{'} = \frac{a-\mu_c}{\sigma_c}$, where μ_c and σ_c are the mean and standard deviation within the corresponding product category c. This problem does not arise in case of the customer attributes since they are relatively comparable. The matrix cell values (# units purchased) could also be very different across categories and have to be standardized. The cell values z_{ij} within each sub-matrix of all the products belonging to a specific product category c are transformed to $z_{ij}^{'} = \frac{z_{ij}-\mu_{z_c}}{\sigma_{z_c}}$ where μ_{z_c} and σ_{z_c} are the mean and standard deviation of the all the values in the sub-matrix.

Algorithm	Training Err.	Test Err.	Test Err. Original
Global Model (k=1,l=1)	0.913 (0.001)	0.920 (0.01)	4.24 (0.06)
Cluster Models (k=4)	0.91 (0.001)	0.917 (0.01)	4.228 (0.059)
CC (k=4,l=4)	0.833 (0.001)	0.890 (0.009)	4.002 (0.056)
Reduced Model (k=4,l=4)	0.849 (0.001)	**0.872 (0.009)**	**3.893 (0.052)**
Model CC (k=4,l=4)	0.804 (0.001)	0.883 (0.007)	3.965 (0.044)

Table 1. Prediction error on the ERIM dataset.

6.1 Predicting Unknown Data Values

Our aim here is to use given purchase information for a set of customers and products along with customer and product attributes to predict the number of units of the product purchased for unknown customer-product combinations. We compare the prediction accuracy of the simultaneous co-clustering and regression approach (Model CC) with a single regression model (Global Model). We also compare Model CC with the special case of the Bregman co-clustering algorithm (CC), which uses squared Euclidean distance and tries to find uniform co-clusters that try to minimize the distance of the data points within the co-cluster to the co-cluster mean [14]. In order use co-clustering (CC) to predict missing values, if missing value z_{ij} is assigned to row cluster g and column cluster h, we estimate z_{ij} with the co-cluster mean μ_{gh}. Additionally, we also compare Model CC with first clustering the customers based on their attributes and then fitting regression models in each customer cluster (Cluster Models).

The prediction error of the different approaches on this problem is computed by averaging over 10 random 90-10% training and test data splits. Table 1 shows the training mean squared error (Training Err.) and the test set error (Test Err.) on the standardized data as well as the test set error on the original, unstandardized dataset obtained by back transforming the standardized data (Test Err. Original) for the different approaches. Model CC and Reduced Model do significantly better than Global Model and Cluster Models on the test set. Model CC and Reduced Model also do slightly better than CC.

Our complete model for the prediction problem consists of the model constructed for the bulk of the matrix entries as described above and a linear model for the high valued outliers. A classifier is trained to appropriately select one of the two models to predict each unknown matrix value. An alternative way of dealing with outliers is to reduce the influence of the extremely high values on the constructed models, allowing them to generalize better. This is achieved by giving high valued matrix entries a very small fixed weight, enabling the linear models to focus on the bulk of the values and rather than fit a few high values. Through experiments conducted for both these cases [10], we observe that here as well, Model CC and Reduced Model do better than the other approaches. Through our experiments on the original dataset, without treating outliers differently [10], we observe that Model CC is very susceptible to overfitting since it is the most complex and involves the most number of parameters. However, Reduced Model performs the best even in this scenario since it has fewer parameters as compared to Model CC and a simpler overall model, that is not as badly affected by outliers.

6.2 Market Segmentation and Structure

Another application of our approach in this domain is that of simultaneous segmentation of customers and products. For this application, the aim is to identify customer segments and groups of "equivalent" products. The simultaneous co-clustering and regression algorithm is used to cluster the customers and products and obtain an interpretable model for each co-cluster, which indicates the factors that influence the purchase decisions of the corresponding subset of customers with respect to the subset of products. Note that the customer and product segments identified by our approach need not be homogenous in terms of customer and product attributes, but are formed in such a way that the attributes have the same coefficients in the corresponding models. Hence, all the customers in a co-cluster will be influenced to a similar degree by the customer and product attributes while making a purchase decision regarding the subset of the products.

The details of our clustering results with 4 customer and 4 product clusters (a total of 16 co-clusters) are illustrated here. Table 2 displays the average number of units purchased in each of the 16 co-clusters. Table 3 shows the linear regression coefficients for 3 interesting co-clusters (Cust. Seg. 3-Prod. Seg. 3, Cust. Seg. 4-Prod. Seg. 4, Cust. Seg. 1-Prod. Seg. 2). This table also compares the co-cluster coefficients with the coefficients of a global model. The values in parentheses are the p-values from the significance tests of the coefficients. One can observe that the coefficients of the sample co-clusters presented here are significantly different from those of the global model, validating our intuition that a single model is inadequate for capturing the heterogeneity in the population. By comparing the regression coefficients across all the co-cluster models [2], we observed that they differ quite a bit across the co-clusters, indicating that different factors are important for different customer and product subsets.

	Prod. Seg. 1	Prod. Seg. 2	Prod. Seg. 3	Prod. Seg. 4
Cust. Seg. 1	1.379	1.378	7.891	0.409
Cust. Seg. 2	0.884	0.650	2.503	0.317
Cust. Seg. 3	1.061	1.197	4.934	1.454
Cust. Seg. 4	1.400	1.480	4.741	0.483

Table 2. Mean # of units purchased in each co-cluster.

Product cluster 3 is a small cluster consisting of the cheapest, most popular and most advertised products. As expected the average number of units purchased is highest for this product cluster across all customer segments (Table 2). All the co-clusters that include product cluster 3, e.g. Cust. Seg. 3, Prod. Seg. 3 as observed in Table 3 have models with very high negative coefficients for price and high positive coefficients for market share and advertising (statistically significant with low p-values), correctly indicating that these are the most influential factors for product preference in these co-clusters. On the other extreme,

[2] The entire set of coefficients along with details of the customer and product clusters identified is available at `http://users.ece.utexas.edu/~deodhar/ERIM`

product cluster 4 consists of products with below average pricing and very low market share and advertising. Overall, across all customer clusters, the average number of units purchased is among the lowest for this product segment. In all the co-cluster models corresponding to this product cluster e.g. Cust. Seg. 4, Prod. Seg. 4, the coefficient for market share is the highest, indicating that this is a major factor influencing the purchase of these products, which is intuitive given the really low market share of the products in this segment. Co-cluster Cust. Seg. 1, Prod. Seg. 2 (Table 3) consists of the customer group with the highest average income and slightly expensive, not well advertised products with average market share. It is interesting that the coefficient for price is positive, indicating that these high income customers, who might not consider price to be an important factor in product choice, don't mind paying more for the products. What drives the sale of the products in this co-cluster is probably their market share and advertising and the large number of shopping visits and amount spent, which end up being very explanatory predictors in the corresponding regression model as seen in Table 3.

Another observation we make based on the co-cluster model coefficients is that on the whole, product attributes are more indicative of customer choice as compared to the customer attributes. Among customer attributes, the total amount spent is a dominating explanatory variable in almost all the regression models. The rest of the customer attributes are not as predictive, probably because they do not vary as much across customer segments and the segment averages for these attributes are close to the global averages. The statistically insignificant predictors for each model can be ignored to get sparse, simpler models.

Attribute	Global	Cust. Seg. 3, Prod. Seg. 3	Cust. Seg. 4, Prod. Seg. 4	Cust. Seg. 1, Prod. Seg. 2
intercept	0.000 (1.000)	-0.423 (0.001)	-0.136 (0.000)	0.151 (0.000)
income	-0.021 (0.000)	-0.089 (0.313)	-0.025 (0.000)	-0.019 (0.424)
# members	0.027 (0.000)	0.028 (0.736)	0.039 (0.000)	-0.042 (0.064)
male head emp.	0.002 (0.424)	-0.060 (0.418)	-0.001 (0.870)	0.052 (0.040)
female head emp.	0.001 (0.617)	-0.071 (0.305)	-0.003 (0.469)	0.016 (0.464)
# shopping visits	0.024 (0.000)	-0.106 (0.054)	0.010 (0.056)	0.112 (0.000)
total amount spent	0.097 (0.000)	0.478 (0.000)	0.029 (0.000)	0.094 (0.000)
price	-0.020 (0.000)	-0.746 (0.000)	-0.022 (0.000)	0.424 (0.000)
market share	0.174 (0.000)	0.425 (0.000)	0.089 (0.000)	0.162 (0.000)
# times advertised	0.096 (0.000)	0.481 (0.000)	0.037 (0.000)	0.041 (0.066)

Table 3. Coefficients of the global model and of 3 co-cluster models.

7 Concluding Remarks

In several marketing applications, due to the heterogeneity of the data it is advantageous to learn segmentwise prediction models rather than a global model. Our model based co-clustering algorithm simultaneously clusters the data along two (or more) modes and builds prediction models, which is better then a sequential approach that independently clusters the data *a priori* and then builds models. Our approach is illustrated on 2 marketing applications, prediction of customer-product preference and segmentation, using the ERIM dataset. The results in section 6 indicate that this approach not only improves prediction accuracy over existing approaches but also produces easily interpretable and actionable models in a segmentation setting. Model based co-clustering can be extended to a wide range of classification, regression and clustering problems in the marketing domain. This framework is not limited to linear regression models and can be easily modified to alternative prediction models that more closely conform to the data characteristics. It would be interesting to explore non-linear models or robust error functions, which would help overcome the limitations of linear models in the presence of outliers.

References

1. Djukanovic, M., Babic, B., Sobajic, D., Pao, Y.: Unsupervised/supervised learning concept for 24-house load forecasting. IEE Proceedings-Generation, Transmission and Distribution **140** (1993) 311–318
2. Lokmic, L., Smith, K.A.: Cash flow forecasting using supervised and unsupervised neural networks. IJCNN **06** (2000) 6343
3. Jordan, M., Jacobs, R., Nowlan, S.J., Hinton, G.E.: Adaptive mixtures of local experts. Neural Computation **3** (1991) 79–87
4. Herlocker, J., Konstan, J., Borchers, A., Riedl, J.: An algorithmic framework for performing collaborative filtering. In: SIGIR. (1999) 230–237
5. George, T., Merugu, S.: A scalable collaborative filtering framework based on co-clustering. In: ICDM. (2005) 625 – 628
6. Madeira, S.C., Oliveira, A.L.: Biclustering algorithms for biological data analysis: A survey. TCBB **1**(1) (2004) 24–45
7. Grover, R., Srinivasan, V.: A simultaneous approach to market segmentation and market structuring. JMR (1987) 139 – 153
8. Wedel, M., Steenkamp, J.: A clusterwise regression method for simultaneous fuzzy market structuring and benefit segmentation. JMR (1991) 385 – 396
9. Moe, W., Fader, P.: Modeling hedonic portfolio products: A joint segmentation analysis of music compact disc sales. JMR (2001) 376 – 385
10. Deodhar, M., Ghosh, J.: A framework for simultaneous co-clustering and learning from complex data. Dept of Elec. and Comp. Engg., Univ. of Texas at Austin, IDEAL-2007-08 (2007) Downloadable from `http://www.lans.ece.utexas.edu/papers/techreports/deodhar07Coclust.pdf`
11. Kim, B., Rossi, P.: Purchase frequency, sample selection, and price sensitivity: The heavy-user bias. Marketing Letters (1994) 57 – 67
12. Kim, B., Sullivan, M.: The effect of parent brand experience on line extension trial and repeat purchase. Marketing Letters (1998) 181 – 193
13. Seetharaman, P., Ainslie, A., Chintagunta, P.: Investigating household state dependence effects across categories. JMR (1999) 488 – 500
14. Banerjee, A., Dhillon, I., Ghosh, J., Merugu, S., Modha, D.: A generalized maximum entropy to bregman co-clustering and matrix approximation. To appear in JMLR (2007) Downloadable from `http://pegasus.ece.utexas.edu/~ghosh/jmlr-coclust.pdf`