Regensburger
DISKUSSIONSBEITRÄGE
zur Wirtschaftswissenschaft

University of Regensburg Working Papers in Business, Economics and Management Information Systems

Hidden Variable Models for Market Basket Data. Statistical Performance and Managerial Implications

Harald Hruschka*

December 15, 2016

Nr. 489

JEL Classification: *C38 ; C45 ; C55 ; M30 ; L81*

Key Words: Marketing; Market Basket Analysis; Factor Analysis; Topic Models; Restricted Boltzmann Machine; Deep Belief Net

* Marketing Chair, Faculty of Business, Economics and Management Information Systems, University of Regensburg, 93040 Regensburg, Germany, Phone: +49-941-943-2279, E-mail: harald.hruschka@wiwi.uni-r.de

Hidden Variable Models for Market Basket Data. Statistical Performance and Managerial Implications

Harald Hruschka

Abstract

We compare the performance of several hidden variable models, namely binary factor analysis, topic models (latent Dirichlet allocation, correlated topic model), the restricted Boltzmann machine and the deep belief net. We shortly present these models and outline their estimation. Performance is measured by log likelihood values of these models for a holdout data set of market baskets. For each model we estimate and evaluate variants with increasing numbers of hidden variables. Binary factor analysis vastly outperforms topic models. The restricted Boltzmann machine and the deep belief net on the other hand attain a similar performance advantage over binary factor analysis. For each model we interpret the relationships between the most important hidden variables and observed category purchases. To demonstrate managerial implications we compute relative basket size increase due to promoting each category for the better performing models. Recommendations based on the restricted Boltzmann machine and the deep belief net not only have lower uncertainty due to their statistical performance, they also have more managerial appeal than those derived for binary factor analysis. The impressive performances of the restricted Boltzmann machine and the deep belief net suggest to continue research by extending these models, e.g., by including marketing variables as predictors.

1 Introduction

We investigate whether and to what extent different hidden variable models are capable to reproduce purchase incidence data for an assortment of several product categories. Purchase incidences are represented by a binary vector. Elements of this vector are set to one if the respective product category is purchased at a purchase occasion. The market basket corresponding to such an purchase incidence vector equals the set of product categories purchased.

We consider several types of hidden variable models, namely binary factor analysis (BFA), two topic models (latent Dirichlet allocation and the correlated topic model), the restricted Boltzmann machine (RBM), and the deep belief net (DBN). Hidden variables identified by these models can be seen as dimensions underlying observed purchases of product categories. These models also provide estimates of how closely each category is associated with a given hidden variable.

In the following we briefly discuss existing publications in which these and related models have been applied to market basket data (broad overviews of methods for market basket analysis can found in Boztuğ and Silberhorn [8] or Reutterer et al. [23]). As early as 1975 Böcker compressed (metric) expenditures of individual customers in eight categories of household goods to four dimension by means of principal component analysis [7]. But to our knowledge purchase incidence data have not been analyzed by appropriate BFA methods though they were already available in the 1980s [2]. The closest study seems to be one of Kamakura and Wedel who introduce a Tobit factor analytic model for mixed metric and binary data [20]. These authors look a two variables (total yearly purchase, purchase incidence) for 39 paper good brands and arrive at a three factor solution.

The use of topic models is widespread in text mining. As a rule they serve to replace words appearing in documents by a relatively small number of discrete hidden variables which are called topics. Quite recently latent activities and latent motivations of customers have been derived by topic models processing purchase incidence data. Hruschka infers ten latent activities based on purchase incidence data for 60 categories offered by a German supermarket [18]. Jacobs et al. reduce 394 categories of chemist's products of a Dutch online retailer to 13 topics [19].

The RBM is frequently applied to solve pattern recognition problems, e.g., recognition of handwritten digits or classification of documents [15]. We are aware of one application to purchase incidences by Hruschka who chooses a RBM with four hidden variables for data on 60 grocery categories [17]. This RBM outperforms a multivariate logit model with category constants and pairwise interactions in a holdout data set.

Deep learning methods to which the DBN belongs turn out to be very successful in different machine learning applications, including speech recognition, computer vision, and natural language processing [22]. In this paper we estimate several variants of a DBN which comprises two hierarchically connected RBMs.

We characterize the contribution of this paper over existing work as follows. The paper provides the first empirical comparison of the performance of several different hidden variables models for multicategory purchase incidence data. It also the first time that binary factor analysis and a deep learning method are applied to such data. In addition to performance measurement we also derive and discuss managerial implications that we obtain on the basis of the better performing methods.

The next section presents the investigated models. We continue by outlining the methods used to estimate these models by emphasizing methods for the RBM and the DBN. The empirical part characterizes the data set, compares model performances for a holdout data set and interprets hidden variables. In the next section we discuss managerial implications. The paper concludes with a summary and an outlook at model extensions which look promising.

2 Investigated Models

2.1 Basic Concepts and Notation

We evaluate models by the log likelihood LL for a holdout data set:

$$LL = \sum_{n=1}^{N} \sum_{j=1}^{J} [y_{nj}\, log(P_{nj}) + (1 - y_{nj})\, log(1 - P_{nj})] \tag{1}$$

N denotes the number of baskets in the holdout data set, J the number of product categories. y_{nj} is a binary purchase incidence which equals one if basket n contains category j. P_{nj} is a shorthand notation for the purchase probability of category j in basket n estimated by a model conditional on hidden variables.

Purchases are collected in a binary purchase indicence vector $y_n = (y_{n1}, \cdots, y_{nJ})$ for each basket n. σ denotes the binary logistic function $\sigma(z) = 1/(1 + exp(-z))$ with real-numbered argument z.

2.2 Binary Factor Analysis

In the binary factor analytic models which we estimate the purchase probability for category j conditional on a vector of hidden variables h_n is specified as:

$$P(y_{nj} = 1 | h_n) = \sigma(b_j + \sum_{k=1}^{K} W_{jk}\, h_{nk}) \tag{2}$$

The conditional purchase probability of category j increases (decreases) monotonically with hidden variable h_{nk} if weight W_{jk} is positive (negative). The K-dimensional vector of hidden variables h_n follows a multivariate normal distribution with zero expectations and covariance matrix Σ.

Though these models allow interdependence of marginal distributions, they specify purchases of a category to be independent of purchases of the other categories conditional on hidden variables (for more details see [1]).

2.3 Topic Models

We consider latent Dirichlet allocation (LDA) and the correlated topic model (CTM) which are the two most frequently used topic models (for a comprehensive review see [28]). To attain compatibility with the other investigated models, we conceive topics as binary hidden variables. In topic models each basket is associated with a topic-mixing vector and each category is independently sampled according to a topic drawn from this mixing vector. In other words, topic models include two multinomial distributions, topic proportions of categories ϕ_{jk} and topic proportions θ_{kn} of each basket n [27].

LDA and CTM differ by the way topic proportions of categories are generated. In LDA topic proportions of categories are generated by a Dirichlet distribution, in CTM by a logistic normal distribution [6]. That is why the CTM allows for correlation between topics.

The probability that basket n contains category j is related to the topic proportions of this category and the proportion of topics of this basket for each topic $k = 1, \cdots, K$ in the following manner [11]:

$$P(y_{nj} = 1|h_n) = \sum_{k=1}^{K} \phi_{jk}\theta_{kn} \tag{3}$$

2.4 Restricted Boltzmann Machines

The restricted Boltzmann machines (RBM) is defined as joint Boltzmann distribution of hidden and observed variables (category purchases) and was introduced by Smolensky [26]. It consists of one layer of observed variables and one layer of binary hidden variables. The RBM is called restricted because variables of the same layer are not connected. Coefficients W_{jk} ($j = 1, \cdots, J$ and $k = 1, \cdots, K$) of a RBM link each observed variable to each hidden variable.

The conditional distribution of each category purchase factorizes given hidden variables and the conditional distribution of each hidden variable factorizes given purchases:

$$\begin{aligned} P(y_{nj} = 1|h_n) &= \sigma(b_j + \sum_{k=1}^{K} W_{jk}\, h_{nk}) \\ P(h_{nk} = 1|y_n) &= \sigma(d_k + \sum_{j=1}^{J} W_{jk} y_{nj}) \end{aligned} \tag{4}$$

b_j and d_k are constants which are specific to category j and hidden variable k, respectively.

We follow Hinton in characterizing the working of a RBM [14]. The probability of a market basket is proportional to the product of the probabilities that the basket would be generated by each of the hidden variables acting alone. If a hidden variable is zero, its separable probability distribution for each category is determined by its category constants b_j only. But if a hidden variable is one, this distribution also depends on the coefficients linking the hidden variable to each category.

In a RBM distributions each specific to a hidden variable are multiplied first. The product of these distributions is normalized in the next step. This way sharp distributions may be detected. Mixture models (to which topic models belong) on the other hand determine convex combinations of distributions which are normalized beforehand. For high dimensional data the mixture model approach may lead to problems, as the final distribution cannot be sharper than the distributions of the individual hidden variables each of which is adapted to all observed variables [14].

The hidden variables of a RBM produce K different partitions of the input space which define 2^K possible regions [3]. Le Roux and Bengio prove that the RBM can approximate any discrete distribution [21]. Therefore the RBM is not restricted to pairwise interactions, but is capable to also reproduce higher order interactions.

2.5 Deep Belief Nets

A deep belief net (DBN) comprises several stacked RBMs. In this paper we consider DBNs with three layers. The first two layers determine hidden variables. The first layer is a RBM based on observed variables (category purchases) corresponding to the RBM presented in section 2.4. Hidden variables of the first layer RBM are used as input variables by the second layer RBM. Based on these input variables the second layer RBM provides second layer hidden variables.

Conditional distributions of inputs and hidden variables for the first and second layers can be written as:

$$\begin{aligned}
P(y_{nj} = 1|h_{1n}) &= \sigma(b_{1j} + \sum_{k=1}^{K} W_{1jk}\, h_{1nk}) \\
P(h_{1nk}|y_n) &= \sigma(d_{1k} + \sum_{j=1}^{J} W_{1jk} y_{nj}) \\
P(h_{1nk} = 1|h_{2n}) &= \sigma(b_{2k} + \sum_{l=1}^{L} W_{2kl}\, h_{2nl}) \\
P(h_{2nl}|h_{1n}) &= \sigma(d_{2l} + \sum_{k=1}^{K} W_{2kl} h_{1nk})
\end{aligned} \tag{5}$$

K, L are the numbers of hidden variables in the first and second layer, respectively. h_{1nk} denotes the kth first layer hidden variable for basket n, h_{2nl} the lth second layer hidden variable of basket n, h_{1v}, h_{2n} are vectors of first and second layer hidden variables for basket n. W_{1jk} denotes the coefficient linking category j to the kth first layer hidden variable, W_{2kl} the coefficient linking the kth first layer hidden variable to the lth second layer hidden variable. b_{1j}, b_{2k}, d_{1k}, d_{2l} are constant terms.

As we want to reconstruct the observed market baskets we add a third layer which directly connects second layer hidden variables to category purchases:

$$P(y_{nj} = 1|h_{2n}) = \sigma(b_{3j} + \sum_{l=1}^{L} W_{3lj}\, h_{2nl}) \tag{6}$$

W_{3lj} denotes the coefficient linking the lth second layer hidden variable to category j, b_{3j} is a category constant.

The two RBMs of the DBN produce nested nonlinear transformations of observed market baskets. Therefore the DBN may provide more abstract representations than a single RBM [5].

3 Estimation

In this section we outline how the different model types introduced in section 2 are estimated. We briefly describe estimation of BFA and topic models and give more details with respect to the RBM and the DBN.

BFA models are estimated by means of the R package *mirt* using a Metropolis-Hastings Robbins-Monro (MH-RM) algorithm [10]. This estimation algorithm starts from random initial values for all coefficients and runs several burn in iterations. Each iteration consists of two steps. In the first step hidden variables are computed by means of Metropolis-Hastings sampling. The second step updates coefficients by a single Newton-Raphson correction for a complete-data gradient vector and Hessian matrix based on the current hidden variables. After burn in final coefficients are determined by the Robbins-Monro root finding algorithm (for more details see [10] and the references given there).

LDA and CTM are estimated by an appropriate variational expectation-maximization (VEM) algorithm implemented in the R package *topic models* [12]. VEM is related to the class of expectation-maximization (EM) algorithms, but replaces the expected complete likelihood which is computationally intractable for both topic models by a variational distribution. The VEM algorithms of *topic models* in fact minimize the Kullback-Leibler (KL) divergence between the variational posterior probability and the true posterior probability. Estimates of hidden variables are determined on the basis of the variational posterior probability (for more details see [12] and the references given there).

Both RBMs and DBNs are estimated using R package *deepnet* [24]. The joint likelihood of the RBM $P(y_n, h_n)$ is related to the so called energy function $E(y_n, h_n)$ with Z denoting the normalization constant [5]:

$$\begin{aligned} P(y_n, h_n) &= \frac{1}{Z} \exp(-E(y_n, h_n)) \\ E(y_n, h_n) &= -\Big(\sum_{j=1}^{J} \sum_{k=1}^{K} W_{jk} y_{nj} h_{nk} + \sum_{j=1}^{J} b_j y_{nj} + \sum_{k=1}^{K} d_k h_{nk} \Big) \end{aligned} \tag{7}$$

The as a rule huge number of configurations of visible and hidden variables prevents direct maximization of the marginal log likelihood. *deepnet* uses for estimation the contrastive divergence (CD) algorithm of Hinton [14] which approximates the marginal log likelihod. The objective of CD is related to the KL divergence between the data distribution and the model distribution which in theory can be produced by infinite many Gibb sampling steps. CD uses only a finite number of T sampling steps instead and determines gradients of RBM parameters as follows:

$$\begin{aligned} & gr(W_{jk}) = \langle y_j h_k \rangle - \langle y_j^T h_k^T \rangle, \\ & gr(b_j) = \langle y_j \rangle - \langle y_j^T \rangle, \; gr(d_k) = \langle h_k \rangle - \langle h_k^T \rangle \end{aligned} \tag{8}$$

$\langle . \rangle$ symbolizes expectations of the product of variables (of the variable) enclosed. N-element vectors y_j and h_k contain purchases of category j and values of hidden variable k, respectively. y_j^T and h_k^T are analogous vectors whose values are generated by T Gibbs sampling steps from the conditional distributions. Step t samples h_n^t given y_n^{t-1} and y_n^t given h_n^t for $t = 1, \cdots, T$ with y_n^0 equal to the observed y_n.

Estimation of the DBN defined in section 2.5 starts with the greedy layerwise algorithm of Hinton et al. which determines parameters of successive RBMs by CD [16]. After estimation of the first layer RBM is finished, its hidden variables are drawn by Gibbs sampling and used as input variables of the second layer RBM.

After estimation of the second layer RBM all hidden variables are set to mean field expected values

μ_{1nk} and μ_{2nl} which are defined as [16]:

$$\begin{aligned} \mu_{1nk} &= \sigma(d_{1k} + \sum_{j=1}^{J} W_{1jk}\, y_{nj}) \\ \mu_{2nl} &= \sigma(d_{2l} + \sum_{k=1}^{K} W_{2kl}\, \mu_{1nk}) \end{aligned} \tag{9}$$

Finally, all parameters of the DBN including those of the third layer are simultaneously estimated by nonlinear least squares. To this end the parameter values of the two RBMs computed by the greedy layerwise algorithm are used as initial values.

4 Empirical Study

4.1 Data

We analyze a publicly available data set which contains one month (30 days) of real-world point-of-sale transactions from a typical local grocery outlet [13]. The data set consists of 9,835 market baskets refering to 169 product categories. Relative frequencies of the 20 most frequently purchased categories are shown in table 1.

Table 1: *Relative Purchase Frequencies*

whole milk	0.256	other vegetables	0.193
rolls/buns	0.184	soda	0.174
yogurt	0.140	bottled water	0.111
root vegetables	0.109	tropical fruit	0.105
shopping bags	0.099	sausage	0.094
pastry	0.089	citrus fruit	0.083
bottled beer	0.081	newspapers	0.080
canned beer	0.078	pip fruit	0.076
fruit/vegetables juices	0.072	whipped/sour cream	0.072
brown bread	0.065	domestic eggs	0.063

lists the 20 highest frequency categories

4.2 Estimation Results

We randomly split the data set into two halves, estimate each model on one half and use the other half as holdout data for which we compute the log likelihood according to expression (1). We evaluate models based on their log likelihood value for the holdout data.

BFA models and the two topic models attain their best log likelihood values for five and six hidden variables, respectively. Among the two topic models the CTM performs better than LDA, but both topic models are vastly inferior to BFA models with less than seven hidden variables. The log likelihood value of the best topic model is lower by more than 25,000 compared to the best BFA model (see table 2). That is why in the following we do not present estimation results of topic models in more detail.

Table 2: *Holdout Log Likelihood Values of BFA and Topic Models*

K	BFA	LDA	CTM
1	-76,576.48		
2	-76,833.23	-99,542.95	-99,543.58
3	-75,811.78	-99,530.46	-99,495.59
4	-75,806.82	-99,548.71	-99,480.04
5	-73,755.72	-99,549.63	-99,456.21
6	-78,363.77	-99,536.24	-99,436.43
7	-91,862.33	-99,558.74	-99,473.76

Table 3 gives holdout log likelihood values for RBMs and DBNs with different numbers of hidden variables. We only show DBNs whose numbers of variables are equal in the first and the second second layers (i.e., $L = K$), because both lower and higher numbers of second layer variables lead to lower log likelihood values.

RBM and DBN start to beat the best BFA model at four and five hidden variables, respectively. The best RBM has 17 hidden variables. Its holdout log likelihood is higher by more than 23,000 compared to the best BFA model. If the DBN has more than nine variables in the second hidden layer it performs better than the RBM with the same number of variables in the first hidden layer. The best DBN has 17 hidden variables both at the first and the second hidden layer. It performs better than the best RBM with a log likelihood higher by about 727.

In the following we interpret the best performing BFA models, RBM and DBN. To this end we look at the importance of each variable in the last hidden layer which we measure by the average sum of absolute marginal effects with respect to purchase probabilities across all 169 categories. This average is computed across all market baskets. Though as a rule the parameter values of a BFA model and a RBM differ, we obtain the same expression for these two models:

$$\frac{1}{N}\sum_n \sum_j P(y_{nj} = 1|h_n)\,(1 - P(y_{nj} = 1|h_n))\,|W_{jk}| \tag{10}$$

For the DBN the analogous expression for a second hidden layer variable l is:

$$\frac{1}{N}\sum_n \sum_j P(y_{nj} = 1|h_{2n})\,(1 - P(y_{nj} = 1|h_{2n}))\,|W_{3lj}| \tag{11}$$

Table 3: *Holdout Log Likelihood Values of RBM and DBN*

K	RBM	DBN
1	-79,276.53	
2	-76,968.97	-77,223.20
3	-74,885.79	-74,584.60
4	-72,147.19	-75,229.42
5	-70,225.06	-71,859.35
6	-68,251.97	-71,116.36
7	-66,620.47	-66,456.30
8	-64,639.61	-65,326.33
9	-61,599.86	-64,234.30
10	-60,209.00	-60,135.23
11	-60,883.98	-57,409.03
12	-57,940.39	-56,516.25
13	-57,792.38	-56,453.84
14	-55,954.81	-54,247.80
15	-53,668.96	- 53,596.10
16	-52,101.18	-51,041.92
17	-50,482.97	-49,755.63
18	-50,504.66	-50,631.04

for all DBN $L = K$

We also assess the importance of a product category j for a hidden variable $k^{'}$ by its information averaged across baskets. Information measures how precise a hidden variable can be estimated using the respective observable variable (here: purchase of a category). It is defined as squared first derivative of the probability of the respective observed variable divided by its variance [9].

For a BFA model and a RBM we average information across baskets and obtain:

$$\frac{1}{N}\sum_{n} P(y_{nj}=1|h_n)\,(1-P(y_{nj}=1|h_n))\,\Big[\sum_{k} W_{jk}corr(k^{'},k)\Big]^2 \tag{12}$$

$cor(k^{'},k)$ denotes the product-moment correlation between hidden variables $k^{'}$ and k.

For a DBN the analogous expression of average information of a product category j with respect to a second layer hidden variable $l^{'}$ is:

$$\frac{1}{N}\sum_{n} P(y_{nj}=1|h_{2n})\,(1-P(y_{nj}=1|h_{2n}))\,\Big[\sum_{l} W_{3lj}corr(l^{'},l)\Big]^2 \tag{13}$$

Note that in the following we simply write sum of absolute marginal effects and information, when we mean their respective averages across market baskets. Table 4 shows for the best BFA model with five hidden variables the sum of absolute marginal effects of hidden variables and informations of category purchases with respect to each hidden variable which are at least 0.15. Hidden variable 2 is missing as all its informations are lower. We obtain higher values of hidden variable 1 for canned beer purchases, but lower values for purchases of dairy products or vegetables/fruits. Hidden variable 3 is similar to hidden variable 1, but in addition ham purchases decrease the value of this hidden variable. Hidden variable 4 is also similar to hidden variable 1 not associated with, e.g., canned beer purchases. Hidden variable 5 differs from the other hidden variables, as it increases for purchases of other vegetables, tropical fruit, or yogurt, but decreases for purchases of canned beer or whole milk.

Tables 5 and 6 list hidden variables whose sum of absolute marginal effects amounts to at least 1.0 for the best RBM and the best DBN, respectively. For each hidden variable seven categories with the highest informations are shown provided that they amount to at least 1.0. In line with their superior statistical performance information values of categories obtained for RBM and DBN are much higher than those obtained for the best BFA model. We also note that the categories which are important to hidden variables are more diverse compared to the best BFA model.

We now interpret hidden variables discovered by the best performing RBM based on table 5. Very high informations indicate which categories are very important for several hidden variables (e.g., shopping bags and rolls/buns for hidden variable 2, shopping bags for hidden variable 3, fruit/vegetable juice and root vegetables for hidden variable 4). Hidden variable 1 increases for purchases of a shopping bag and decreases for purchases of specialty fat, flour, pasta, coups, baking powder, or pasta. Hidden variable 2 decreases for purchases of shopping bags, rolls/buns, dish cleaner, flour, canned fish, instant food products, or mayonnaise. Hidden variable 3 increases if shopping bags, white whine, ice cream, or liquor are purchased.This hidden variable decreases for purchases of rolls/buns, bathroom cleaner, or cooking chocolate. Hidden variable 4 is higher if root vegetables or herbs are purchased and lower for purchases of fruit/vegetables juice or specialty bar.

Table 4: *Sum of absolute marginal effects and informations (best BFA model)*

hidden variable 1	2.15		hidden variable 3	1.46	
whole milk	0.30	-	whole milk	0.40	-
other vegetables	0.34	-	canned beer	0.33	+
yogurt	0.27	-	yogurt	0.25	-
root vegetables	0.25	-	other vegetables	0.24	-
tropical fruit	0.24	-	tropical fruit	0.23	-
canned beer	0.24	+	ham	0.15	-
butter	0.16	-			
whipped/sour cream	0.16	-			
hidden variable 4	1.29		hidden variable 5	1.27	
whole milk	0.18	-	canned beer	0.20	-
yogurt	0.18	-	whole milk	0.20	-
other vegetables	0.17	-	other vegetables	0.18	+
tropical fruit	0.17	-	tropical fruit	0.18	+
			yogurt	0.16	+

+ and + indicate sign of W_{jk}; contains information values >= 0.15 only

We continue by interpreting hidden variables of the best performing DBN based on table 6. Very high informations show which categories are very important for several hidden variables (whipped/sour cream for hidden variable 1, pip fruit and citrus fruit for hidden variable 2, tropical fruit for hidden variables 4 and 6, domestic eggs for hidden variable 7, yogurt for hidden variable 8). Hidden variable 1 assumes low values if whipped sour cream, citrus fruits, coffee, butter, pip fruit, sugar, and margarine are purchased. Hidden variable 2 increases for purchases of citrus fruit or rolls/buns, it decreases if pip fruit, shopping bags, ice cream, white wine, or house keeping products are purchased. Hidden variable 3 becomes low for purchases of newspapers, whipped/sour cream, citrus fruit,or oil. Hidden variable 4 increases if canned beer is purchased. It decreases for purchases of tropical fruit, bottled water, grapes, or turkey. Hidden variable 5 increases if butter milk is purchased and decreases if sausages or bottled beer are purchased. Hidden variable 6 is positively associated with purchases of bottled water or canned beer, negatively with purchases of tropical fruit. Hidden variables 7 and 8 assume low values for purchases of domestic eggs or yogurt, respectively.

We do not obtain high product moment correlations in absolute terms between RBM and DBN for the hidden variables listed in tables 5 and 6 except for the two variables 3 and 2 with a correlation coefficient of -0.72. This result can be explained by the different signs of that the coefficients for shopping bags and white wine have for these two hidden variables.

Table 5: *Sum of absolute marginal effects and informations (best RBM)*

hidden variable 1	2.83		hidden variable 2	1.53	
shopping bags	9.39	+	shopping bags	23.67	-
cling film/bags	7.19	-	rolls/buns	11.51	-
specialty fat	6.80	-	dish cleaner	2.40	-
flour	6.64	-	flour	2.12	-
pasta	6.14	-	canned fish	2.04	-
soups	6.10	-	instant food products	1.95	-
baking powder	6.05	-	mayonnaise	1.93	-
hidden variable 3	1.31		hidden variable 4	1.00	
shopping bags	19.60	+	fruit/vegetable juice	37.01	-
rolls/buns	5.84	-	root vegetables	23.67	+
white wine	5.63	+	specialty bar	1.98	-
ice cream	3.28	+	herbs	1.46	+
bathroom cleaner	1.71	-			
liquor	1.30	+			
cooking chocolate	1.28	-			

contains hidden variables with sum of absolute marginal effects ≥ 1.0.

+ and + indicate sign of W_{jk}; contains seven highest information values if > 1.0

Table 6: *Sum of absolute marginal effects and informations (best DBN)*

hidden variable 1	3.50		hidden variable 2	2.25	
whipped/sour cream	22.39	-	pip fruit	50.00	-
citrus fruit	7.40	-	citrus fruit	15.05	+
coffee	6.75	-	shopping bags	5.99	-
butter	5.83	-	rolls/buns	2.83	+
pip fruit	5.57	-	ice cream	2.36	-
sugar	5.28	-	white wine	2.14	-
margarine	5.10	-	house keeping products	1.09	-
hidden variable 3	1.68		hidden variable 4	1.30	
newspapers	4.75	-	tropical fruit	17.78	-
whipped/sour cream	2.43	-	bottled water	6.00	-
citrus fruit	1.29	-	canned beer	2.18	+
oil	1.11	-	grapes	1.53	-
			turkey	1.05	-
hidden variable 5	1.19		hidden variable 6	1.16	
sausage	4.97	-	tropical fruit	13.95	-
bottled beer	4.06	-	bottled water	5.89	+
butter milk	2.51	+	canned beer	3.27	+
hidden variable 7	1.15		hidden variable 8	1.01	
domestic eggs	13.69	-	yogurt	16.45	-

contains hidden variables with sum of absolute marginal effects ≥ 1.0.

+ and + indicate sign of W_{3lj}; contains seven highest information values if > 1.0

5 Managerial Implications

To derive managerial implications of this research we assess the effect of a promotion which increases the purchase probability of a category on an objective which takes all categories into account. As data on category specific average sales revenues or margins are not available, we look at the relative increase of basket size averaged across baskets. Basket size corresponds to the sum of purchase probabilities across all 169 categories, which are estimated by means of the chosen BFA model (five hidden variables), RBM and DBN (both with 17 variables in the last hidden layer). We compute two different basket sizes, bs_0 based on the observed data, and bs_{1j} based on the assumption that the respective category j is added to 500 baskets which are randomly selected from the baskets which do not contain this category. Relative basket size increase is defined as $(bs_{1j} - bs_0)/bs_0$. A manager who wants to increase average basket size should execute promotions in categories with high estimated relative basket size increases

Table 7 lists the 20 categories with highest relative basket size increases for each of the three selected models. The categories with higher relative basket size increases differ extremely between the BFA model on one hand and the RBM and DBN on the other hand. Only two of the 20 categories, root vegetables and whipped/sour cream, given for the BFA model belong to the 20 categories with highest purchase frequencies of table 1. These two categories are also the only ones which can be found in the columns for the RBM and the DBN. Because of its clearly worse statistical performance, the categories given for the BFA model should not be recommended for a promotion whose objective is to increase basket size.

In addition high computation times also suggest not to base such decisions on the BFA model. The computation time to estimate basket size is higher by a factor greater than 100 compared to the RBM and the DBN. This fact is caused by more than 1.6 million five dimensional integrals which must be approximated (one integral for each category and each observed basket).

The 20 categories with the largest relative basket size increase according to the best RBM and DBN include 15 and 17 of the 20 categories with highest purchase frequencies of table 1, respectively. But it is wrong to assume that basket size increases with the purchase frequency of a category, as Spearman correlation coefficients between relative purchase frequencies and relative basket size increases demonstrate. For the RBM there is no monotone relationship between the two rankings as a correlation of -0.05 shows. For the DBN the correlation is -0.70 which means that the two rankings are to some extent reverse to each other. This result agrees with the fact that the three categories with the highest relative basket size increases according to the DBN (pip fruit, whipped/sour cream, domestic eggs) are among the five categories with the lowest purchase frequencies given in table 1.

Candidates for promotions are categories with high estimated relative basket size increases according to the DBN, because it performs best from a statistical point of view. Of course, such decisions must be supplemented by managerial judgment, as coefficients of the models investigated here do not necessarily reflect causal effects. E.g., it can be doubted that promoting shopping bags entails an increase of basket size. It seems more obvious that a higher baskets size causes customers to also buy a shopping bag.

6 Conclusions

We compare the capability of several models with hidden variables to reproduce purchase incidence data for several product categories. As model types we consider two topic models (latent Dirichlet allocation and the correlated topic model), binary factor analysis (BFA), the restricted Boltzmann machine (RBM), and the deep belief net (DBN).

We evaluate variants of these models each with different numbers of hidden variables by their log likelihood for a holdout data set. Topic models turn out to be vastly inferior to the BFA model which on its own is vastly outperformed by the RBM and the DBN. The DBM attains better log likelihood

Table 7: *Relative Basket Size Increases By Promoting a Category*

BFA		RBM		DBN	
preservation products	0.099	shopping bags	0.081	pip fruit	0.105
baby food	0.079	other vegetables	0.076	whipped/sour cream	0.079
baby cosmetics	0.073	domestic eggs	0.072	domestic eggs	0.077
herbs	0.038	fruit/vegetable juice	0.069	citrus fruit	0.077
ready soups	0.036	yogurt	0.068	frankfurter	0.076
pudding powder	0.035	whole milk	0.067	newspapers	0.076
rice	0.035	frankfurter	0.064	yogurt	0.066
root vegetables	0.033	pastry	0.063	bottled beer	0.061
sliced cheese	0.033	root vegetables	0.062	bottled water	0.060
whipped/sour cream	0.031	newspapers	0.059	sausage	0.058
abrasive cleaner	0.031	sausage	0.059	coffee	0.057
cream cheese	0.030	soda	0.059	shopping bags	0.055
butter milk	0.030	rolls/buns	0.055	specialty fat	0.054
flour	0.030	specialty fat	0.053	fruit/vegetable juice	0.051
kitchen utensil	0.029	bottled water	0.053	soda	0.051
onions	0.028	tropical fruit	0.048	pastry	0.049
jam	0.027	dish cleaner	0.046	tropical fruit	0.049
liver loaf	0.027	bottled beer	0.042	other vegetables	0.048
specialty cheese	0.026	specialty vegetables	0.041	whole milk	0.048
curd	0.026	cling film/bags	0.041	root vegetables	0.044

lists 20 categories with highest increases

values than the RBM.

To infer managerial implications we estimate relative basket size increases due to promoting a product category for the best performing BFA model, RBM, and DBN. Product categories with high relative increases constitute candidates for a promotion whose the objective is to increase basket size. Recommendations based on the RBM and the DBN not only have lower uncertainty due to the statistical performance of these models, but to our opinion have more managerial appeal than those derived for the BFA model.

The impressive performance advantages of the RBM and the DBN suggest to continue research by specifying and estimating appropriate extensions of these models. To include predictors, especially marketing variables on price and sales promotion, seems to be an obvious next step. Other possible extensions consist in using other dependent variables, e.g., purchase quantity or expenditure. It might also be feasible to take a more detailed perspective by considering several brands in each of a higher number of product categories. Existing research so far has either remained at the category level, or looked at several brands of one category or considered very few brands in very few categories (see, e.g., the overviews [25] or [8]).

References

[1] Ackerman TA (2005). Multidimensional item response theory models. In Everitt BS, Howell DC (eds.) Encyclopedia of statistics in behavioral science, Volume 3. Wiley, Chichester, pp. 1272–1280

[2] Bartholomew DJ (1980) Factor analysis for categorical data. J Roy Stat Soc B 42:293-321

[3] Bengio Y (2009) Learning deep architectures for AI. Fdn Trends Mach Learn 2:1–27

[4] Bengio Y, Courville A, Vincent P (2013) Representation learning: A review and new perspectives. IEEE Trans Pattern Anal Mach Intell 35:1798-1828

[5] Bengio Y, Lamblin P, Popovic D, Larochelle H (2007) Greedy layer-wise training of deep networks. In Schölkopf B, Platt J, Hoffman, T (eds) Advances in neural information processing systems 19 (NIPS'06). MIT Press, Cambridge, MA, pp. 153–160

[6] Blei DM, Lafferty JA (2007) A correlated topic model of science. Ann Appl Stat 1:1735

[7] Böcker F (1975) Die Analyse des Kaufverbunds -Ein Ansatz zur bedarfsorientierten Warentypologie. Z betriebswirt Forsch 27:290-306

[8] Boztuğ Y, Silberhorn N (2006) Modellierungsansätze in der Warenkorbanalyse im Überblick. J Betriebswirt 56:105–128

[9] Brown A, Croudace T (2015) Scoring and estimating score precision using multidimensional IRT. In Reise SP, Revicki, DA (eds.) Handbook of item response theory modeling: Applications to typical performance assessment. Routledge/Taylor & Francis, New York, pp. 307–333

[10] Chalmers RP (2012) mirt: A multidimensional item response theory package for the R environment. J Stat Softw 48 (6):1-29

[11] Griffiths TL, Steyvers M (2004) Finding scientific topics. Proceedings of the National Academy of Sciences (Suppl. 1) 101:5228–5235

[12] Grün B, Hornik K (2011) "topicmodels: An R package for fitting topic models. J Stat Softw 40 (13):1–30

[13] Hahsler M, Hornik K, Reutterer T (2006) Implications of probabilistic data modeling for mining association rules. In Spiliopoulou M, Kruse R, Borgelt C, Nürnberger A, Gaul W (eds) From data and information analysis to knowledge engineering. Springer, Berlin, pp. 598–605

[14] Hinton GE (2002) Training products of experts by minimizing contrastive divergence. Neural Comp 14:1771–1800

[15] Hinton GE, Salakhutdinov RR (2006) Reducing the dimensionality of data with neural networks. Sci 313:504–507

[16] Hinton GE, Osindero S,Teh Y (2006) A fast learning algorithm for deep belief nets. Neural Comp 18:1527–1554

[17] Hruschka H (2014a) Analyzing market baskets by restricted Boltzmann machines. OR Spectrum, 36:209–228

[18] Hruschka H (2014b) Linking multi-category purchases to latent activities of shoppers: Analysing market baskets by topic models. Market ZFP 36:267–274

[19] Jacobs B, Donkers B, Fok D (2016) Model-based purchase predictions for large assortments. Mark Sci 35: 389-404

[20] Kamakura WA, Wedel M (2001) Exploratory Tobit factor analysis for multivariate censored data. Multiv Behav Res 36:5–82

[21] Le Roux N, Bengio Y (2007) Representational power of restricted Boltzmann machines and deep belief networks. Technical Report 1294, Département d'informatique et recherche opérationnelle, Université de Montréal

[22] Najafabadi MM, Villanustre F, Khoshgoftaar TM, Seliya N, Wald R, Muharemagic E (2015) Deep learning applications and challenges in big data analytics. J Big Data doi:10.1186/s40537-014-0007-7

[23] Reutterer T, Hahsler M, Hornik K (2007) Data Mining und Marketing am Beispiel der explorativen Warenkorbanalyse. Market ZFP 29:163-179

[24] Rong X (2014) deepnet: deep learning toolkit in R. R package version 0.2. https://CRAN.R-project.org/package=deepnet. Accessed 3 November 2016

[25] Seetharaman PB, Siddhartha C, Ainslie A, Boatwright P, Chan T, Gupta S, Mehta N, Rao V, Strijnev A (2005) Models of multi-category choice behavior. Market Letters 16:239–254

[26] Smolensky P (1986) Information processing in dynamical systems: foundations of harmony theory. In: Rumelhart DE, McClelland JL (eds) Parallel distributed processing: explorations in the microstructure of cognition. Vol. 1: foundations. MIT Press, Cambridge, pp. 194–281

[27] Steyvers M, Griffiths T(2007) Probabilistic topic model. In: Landauer T, McNamara D, Dennis S, Kintsch W. (eds) Handbook of latent semantic analysis. Erlbaum, Hillsdale, NJ, pp. 424–440

[28] Sun Y, Deng H, Han J (2012) Probabilistic models for text mining. In Aggarwal CC; Zhai CX (eds) Mining text data. Springer, New York, pp. 259-295