Exploring Customer Preferences with Probabilistic Topics Models

Konstantinos Christidis, Dimitris Apostolou, Gregoris Mentzas

National Technical University of Athens,
Iroon Polytechniou 9 Zografou Athens, 157 80 Greece
{kchrist,dapost,gmentzas}@mail.ntua.gr

Abstract. Customer preference learning and recommendation for (e-)commerce is a widely researched problem where a number of different solutions have been proposed. In this study we propose and implement a novel approach to the problem of extracting and modelling user preferences in commerce using latent topic models. We explore the use of probabilistic topic models on transaction itemsets considering both single one-time actions and customers' shopping history. We conclude that the extracted latent models not only provide insight to the consumer behaviour but also can effectively support an item recommender system.

Keywords: Customer Preferences, Latent Topics, LDA.

1 Introduction

Marketing research has long ago applied data mining and machine learning techniques in retail transactions in which large amounts of purchase data have been analyzed [27], [28]. Market basket analysis discovers association patterns in retail transactions and lays the foundations for applications such as product bundling, cross category dependency identification as well as consumer profiling [36], [10].

The advent of electronic commerce has paved the way for the numerous advances in techniques and models intended for enhancing the customer experience in electronic stores. Among them, preference learning attempts to specify consumer desires in a declarative way in order to, for example, support the recommendation of new products [19]. Typically, a recommender system compares a user profile to some reference characteristics, and seeks to predict the 'rating' that a consumer would give to an item they had not yet considered. These characteristics may be derived from the item itself (the content-based approach) or the user's social environment (the collaborative filtering approach) [1].

Applying recommender systems' techniques to market basket data faces several challenges. First, the commonly used approach of association rules mining provides limited insight in the underlying structure of the user preferences. Even though we can use association rules to successfully predict the remainder of the user's basket, we lack a total view of user tastes and their relations. Second, there are a number of technical issues relating to the most common recommendation techniques.

Association rules tend to ignore large itemsets, and memory-based collaborative learning lacks scalability [12]. On the other hand, content based recommenders are inappropriate since information about retail products is neither readily available nor appropriately detailed.

In an effort to address the aforementioned challenges of existing recommender systems' techniques used in market basket analysis, we explore the use of latent topic models. Latent topic models are statistical models used for analyzing relationships between a set of documents and the terms they contain by producing a set of concepts related to the documents and terms. Latent topic models have been highly successful in applications such as information retrieval and filtering, natural language processing, machine learning from text, and related areas (eg. [3],[20],[24]). In this paper we apply latent topic models in order to provide a model for consumer preferences as well as to effectively recommend products to consumers. In particular, we explore latent topic models to discover latent baskets and latent users from purchase data, we propose a recommendation mechanism based on latent baskets and latent users and we compare results with recommendations derived from association rule mining, a technique typically used in market basket analysis.

This paper is structured as follows. The next section provides a brief overview of the related work, while section three briefly discusses topic models and Latent Dirichlet Allocation in particular. Section four describes the experiment methodology and section five provides an analysis of the results. Subsequently we present results of this study as well as plans for further work.

2 Related Work

Recommender systems are addressing, in the most general expression, the problem of estimating the utility or the ratings of items that have not yet been seen by the user [1]. In web-based e-commerce, information overload has created challenges to customer selection of products for online purchases and to sellers attempting to identify customers' preferences efficiently [17]. Business interest in this area started from early adopting electronic businesses, such as Amazon.com, and is growing since [23]. Various types of recommendation techniques have been researched:

- Content-based, where the content of each item is analysed and mapped against the user's past preferences in order to predict his future ratings.
- Collaborative filtering, where the behaviour of a number of users is analysed and it is assumed that similarly behaving users in the past will continue to behave in a similar way in the future,
- Knowledge engineering, where human effort is needed in order to discover the factors that affect users' preferences [17].
- Case-based recommenders that treat the objects to be recommended as cases and use the recall of examples as the fundamental problem-solving process [7].
- Hybrid recommenders that combine the above methods.

Recommender systems are widely used in e-commerce applications [35]. E-commerce poses specific challenges to recommender systems such as scarcity, scalability and quality of recommendations [33], [11]. Specific systems such as

agent-based systems [22], [13] and techniques such as dimensionality reduction, generative models, spreading activation and link analysis [16] have been proposed for recommending products and services in e-commerce applications. Specifically in [18] the authors research a number of probabilistic models focusing on users and evaluate them in an online download site.

In analysing customer transactions in the so called market basket analysis, early research has focused on the application of association rule mining. Research efforts have incorporated estimation and pruning techniques [2], [34]. Association rule mining involves the analysis of itemsets, and the extraction of rules that relate products. In this direction there has been work selecting only the non-derivable itemsets [8]. In a slightly different direction, collaborative filtering techniques have been researched for use in market basket data [21], [26].

Mined association rules are easy to understand and many new patterns can be identified. However, the sheer number of association rules may make the interpretation of the results difficult [9]. Collaborative filtering also may have decreased performance in large datasets. In this work we explore the application of probabilistic topic models in an effort to provide a model for consumer behaviour that interprets market basket data. Moreover, we aim to provide a means to effectively recommend products to consumers even in large datasets.

3 Probabilistic Topic Models

Basic Assumptions

Probabilistic topic models are generative models that have been developed in order to sufficiently describe document corpora. It is hypothesized that there is a way to generate the documents by using latent topic variables and probabilistic sampling techniques, i.e. documents can be generated by sampling topics from a distribution of topics over documents and sampling words from a distribution of words over each topic. In order to describe the model these distributions need to be identified. Statistical inference is used in order to approximate the underlying model, which is most probable to have generated this data.

Latent Dirichlet Allocation

Latent Semantic Analysis was the first probabilistic topic model technique to analyze documents and the words that they contain in order to generate a set of concepts that relate to both of them. Probabilistic Latent Semantic Analysis, as proposed by Hofmann in [15], is an evolution of the previous model that incorporated a probabilistic foundation. Latent Dirichlet Allocation (LDA) was then proposed, assuming that both word-topic and topic-document distributions have a Dirichlet prior.

LDA is based on probabilistic principles and overcomes problems previously encountered in probabilistic Latent Semantic Analysis. Topic mixture weights are not individually calculated for each document, but are treated as a k-parameter hidden random variable (where k is the number of topics). Therefore the method is naturally generalized to new documents. Additionally the parameters needed are not growing with the size of the training corpus. Being a statistical model, LDA requires training in order to converge to a generative model. Afterwards, this model can be used in applications.

The graphical model in Fig. 1, as found in [5], illustrates in plate notation the generative model: z and d variables identify topics and documents, while θ(d) is the distribution over topics for a document d and φ(z) is the distribution over words for a topic z. These distributions can be used to generate documents in the form of a collection of words (w). D is the number of documents, T is the number of topics in the corpus and Nd the topics found in each document. Hyperparameters α and β identify the Dirichlet priors of the above multinomial distributions respectively. These hyperparemeters can be changed in order to control the smoothing of the distributions.

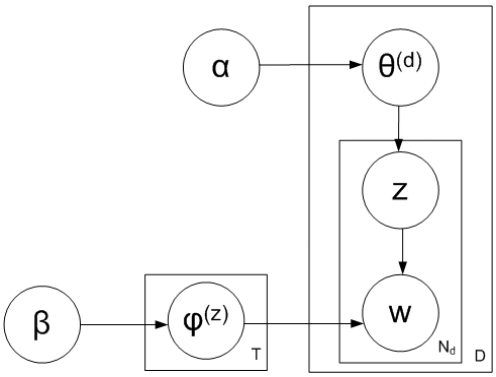

Fig. 1. LDA Plate Notation [5]

After the topics have been trained, it is possible to infer the distribution that could have generated a new, previously unseen, item.

4 Applying Probabilistic Topic Models to Market Baskets

We apply probabilistic topic models and specifically Latent Dirichlet Allocation on retail transaction data. In this work we use a variant of the model that places a Dirichlet prior on both distributions of topics over documents and of words over topics as discussed in [5]. In this section we describe how we extract topic models using both the itemsets contained in transactions and the aggregation of the consumer preferences over time.

Problem Definition

A market basket is composed of items bought together in a single trip to a store. The most significant attributes are the transaction identification and item identification. While ignoring the quantity bought and the price, each transaction represents a purchase, which occurred in a specific time and place. This purchase can be linked to an identified customer (usually carrying a card) or to a non-identified customer. The data set with multiple transactions can be shown in a relational table (transaction, item). The table (transaction, item) is a set of all transactions

$$T = \{T_1, T_2, T_3, ..., T_n\}. \quad \textbf{(1)}$$

where each transaction can be modelled as a binary vector. T is a vector where t[k] =1 if the transaction contained the item I_k, and t[k] =0 otherwise. [9],[30]

$$I = \{I_1, I_2, I_3, ..., I_K\}. \quad \textbf{(2)}$$

Based on the attributes (transaction, item), the market basket is defined as the N items that are bought together more frequently. The next step is to identify all the customers having bought N-m items of the basket and suggest the purchase of some m missing items. In order to make decisions in marketing, the market basket analysis is a powerful tool supporting the implementation of cross-selling strategies.

Model Extraction

In order to apply topic model analysis we need to consider a corpus of documents that consist of terms. The terms in our case are the products available for sale. We consider each document to be formed either by the products bought together in a single transaction or by the products bought by a specific customer.

Latent Baskets

First, we consider each transaction to be a document. The products bought together by a customer during a single trip are considered as a document created using words from the product list. This itemset is then assumed to be a result of a generative topic model which we try to compute. This results into a collection of probabilistic itemsets of products that occur together in documents and can be seen as latent baskets.

Latent Users.

Second, we consider the series of transactions performed by a customer to be a document. The products bought by a customer in the course of time are considered as a document created using words from the product list. This itemset is then assumed to be a result of a generative topic model which we try to describe using topic models.

The topics of these models are considered to reflect the customer taste over the period of time

Deriving Recommendations

Next, we utilise the extracted latent topic models in order to provide recommendations for the users. For each available item, we calculate its similarity (w) to the items already in the user basket. Then the most similar items are recommended.

We employ Gibbs Sampling and a thesaurus based approach comparable to [29] to infer relevant products. We also boost co-occurrence of products in multiple topics and then combine both basket and user topic models.

The following techniques have been tested for recommending items.

1. Latent Baskets — Gibbs Sampler. In this case the latent baskets are used in order to predict the behaviour of a user, given the items he has in his basket right now. A Gibbs sampler is used in order to infer the probability distribution of the known items contained in a basket. Subsequently the items with the highest affinity to this distribution that are not already in the known items list are suggested to the user.
2. Latent Baskets Thesaurus. As in the previous case, the latent baskets are used in order to predict the behaviour of a user. However instead of using statistical inference, a thesaurus is generated during the creation of the topic models. To remove the document dependence from our calculations, we examine term–term relationships instead of term–document relationships. The resulting structure connects words and similarities between them.

Here we consider $S_{LBi,j}$ to be the calculated similarity between items i and j because of the latent baskets topic model. We use equation 3 to compute the similarity of the known items (KI) of the user basket to the n different possible items. Then the top suggestions are compared to the real choices of the user.

$$\{w_1, w_2, w_3, ..., w_n\} = \{\sum_{j \in KI} S_{LB1,j}, \sum_{j \in KI} S_{LB2,j}, \sum_{j \in KI} S_{LB3,j}, ..., \sum_{j \in KI} S_{LBn,j}\}. \quad (3)$$

3. Latent Baskets with Co-occurrence Boosting. This case is similar to using a thesaurus, but in this case items found relevant with more than one known items from the user's basket receive a slight boost. We count the number of similar items known to be in the user's basket and use their number as a power for the co-occurrence factor (b). In (4), M is the number of items in the basket found similar to the item l.

$$w_l = b^{M-1} \sum_{j \in KI} S_{LBl,j} \ . \quad (4)$$

4. Latent Users Thesaurus. In this case we use the topic models extracted by the user's preferences over time. We generate a thesaurus where $S_{LUi,j\ is}$ the calculated similarity between items i and j because of the latent users topic model. We use these similarities in order to predict the suggested items, given the known items in a user's basket as presented in (5)

$$\{w_1, w_2, w_3, ..., w_n\} = \{\sum_{j \in KI} S_{LU1,j}, \sum_{j \in KI} S_{LU2,j}, \sum_{j \in KI} S_{LU3,j}, ..., \sum_{j \in KI} S_{LUn,j}\}. \quad (5)$$

5. Latent Baskets combined with Latent Users. In this case the latent baskets models recommendation is complemented by a latent user model, using a mixing parameter μ, $0<\mu<1$. To compute the similarity of the item l to the known items in a user's basket we use (6)

$$w_l = (1-\mu)\sum_{j \in KI} S_{LBl,j} + \mu \sum_{j \in KI} S_{LUl,j}. \quad (6)$$

Additionally, we use fp-growth [14] association rules mining as a baseline algorithm against which we compare the results of our models. FP-growth is an efficient, fast and scalable method that generates an extended prefix-tree structure called FP-tree and a mining method for this structure.

5 Experiments and Results

Dataset

The transaction data have been collected from a major super market in Europe during a one-year time period. The number of transactions is 1,057,076 while the number of identified customers is 17,672. The different items available for purchase are 102,142. In the process of this experiment we have used product categories in order to draw useful results connecting types of products, without distinguishing different brands. For example, instead of counting different brands and carton sizes of milk, we consider low-fat milk to be a single product. This granularity results to 473 distinct products.

Using this dataset we applied the topic analysis techniques on baskets and users as described above. The results of the analysis can not only be the basis for recommendation but can also provide insight into customers purchasing behaviour. Topic extraction using LDA and the mallet framework [25] took 4 hours and 30 minutes for 2000 iterations in an Intel Core 2 Duo T9300 CPU with 2.0 GB of RAM.

Table 1 presents indicative examples of latent baskets that intuitively reflect groups of products bought together. Table 2 presents indicative examples of latent users mined from the same dataset.

Table 1. Examples of latent baskets.

Latent basket id	Items
23	White paper napkins, Body shampoo, Snack, Soda
25	Toilet paper, Kitchen paper, White paper napkins, Oily hair shampoo
29	Toilet paper, Kitchen paper, White paper napkins, Bleach
34	Spoons/forks/knives, kitchen utensils, daily use, Plastic utensils
40	Cleaning sponges, Cleaning towels, Scourers
42	Olives, Pickles, Precooked food can, Canned fish

Table 2. Examples of latent users.

Latent user id	Items
3	Toothpaste, Cleaning Sponges, Bleach, Water descaler
6	Gouda cheese, Kaser cheese, edam cheese, feta cheese,
7	Brooms, Broom sticks, mops, mop towels
9	School accessories, Ravioli pasta, Fresh milk
24	Hygiene Cotton , baby napkins, Colored napkins
39	Cake bases, truffle, Turkish delights, semolina

Results

In this section we describe the results of the evaluation of the latent topics based recommender.

We used the FP-growth association rule mining in different configurations as a baseline recommendation system [25],[14], using the rapidminer platform [32]. We got 12522 rules from the dataset using 1000 frequent itemsets and a minimum confidence of 0.1. These rules are ranked based on their respective confidence and applied on the known items of each active basket.

In Table 3, we depict the iterations required in order to achieve a convergence to representative topic models. We evaluate the best performance for the recommendation algorithm using thesaurus for inference. We observe that we achieve best results after 4000 iterations, while comparably good results are achieved after as early as 2000 iterations.

Table 3. Evaluating LDA iterations

LDA\iterations	**500**	**1000**	**2000**	**3000**	**4000**
Precision	0.0976	0.1151	0.1322	0.1358	0.1385
Recall	0.3228	0.3835	0.4406	0.4526	0.4618
F- measure	0.1499	0.1771	0.2034	0.2089	0.2131

We then proceed to evaluate the outcome of the recommendations for different values of parameters for Latent Baskets co-occurrence boosting and for combining Latent Baskets with Latent Users. The results can be found in Table 4 and Table 5.

Table 4. Co-occurence boosting factors

Co-occurrence Boosting (b)	**1.1**	**1.3**	**1.5**	**1.7**	**2**
Precision	0.1080	0.1097	0.1139	0.1105	0.1086
Recall	0.1799	0.1842	0.1898	0.1842	0.1810

Table 5. Mixing Parameter for Latent Baskets and Latent Users

Mixing Parameter (μ)	**0.1**	**0.3**	**0.5**	**0.7**	**0.9**
Precision	0.1108	0.1255	0.1135	0.1169	0.1215
Recall	0.1847	0.2155	0.1891	0.1948	0.2025

Using b=1.5 and μ=0.3 we performed experiments with recommendation sizes of 3 to 21 items. All the techniques mentioned before were applied on the 80% of the dataset while the remaining 20% was held for evaluation. A presentation of some of the results is contained in Table 7. The Precision/Recall curves of the different techniques reflect the full results of the experiments and can be found in Fig. 2.

Table 6. Results for all methods and different recommendation sizes

	Recommended Items	**5**	**9**	**13**	**19**
a. Latent Baskets - Gibbs Sampler	**Precision**	0.1476	0.1221	0.1146	0.0930
	Recall	0.2499	0.3662	0.4967	0.5893
b. Latent Baskets Thesaurus	**Precision**	0.1124	0.1140	0.1037	0.0883
	Recall	0.1875	0.3408	0.4475	0.5563
c. Latent Baskets with Co-occurrence Boosting	**Precision**	0.1114	0.1079	0.0976	0.0858
	Recall	0.1856	0.3235	0.4225	0.5130
d. Latent Users Thesaurus	**Precision**	0.1239	0.1101	0.0990	0.0826
	Recall	0.2065	0.3303	0.4289	0.5232
e. Latent Baskets combined with Latent Users	**Precision**	0.1145	0.1152	0.1044	0.0919
	Recall	0.1908	0.3457	0.4524	0.5502
f. FP-Growth Association Rules	**Precision**	0.0436	0.0687	0.0747	0.0763
	Recall	0.0723	0.2000	0.2952	0.3116

The evaluation results are summarized in Fig. 2, where the precision/recall curves of the different techniques are depicted. For the first k recommendations of each technique we evaluate and compare to the real items bought and we compute precision and recall. For different values of k, we trace the precision/recall curve for each method revealing how accurate and how complete our recommendations are.

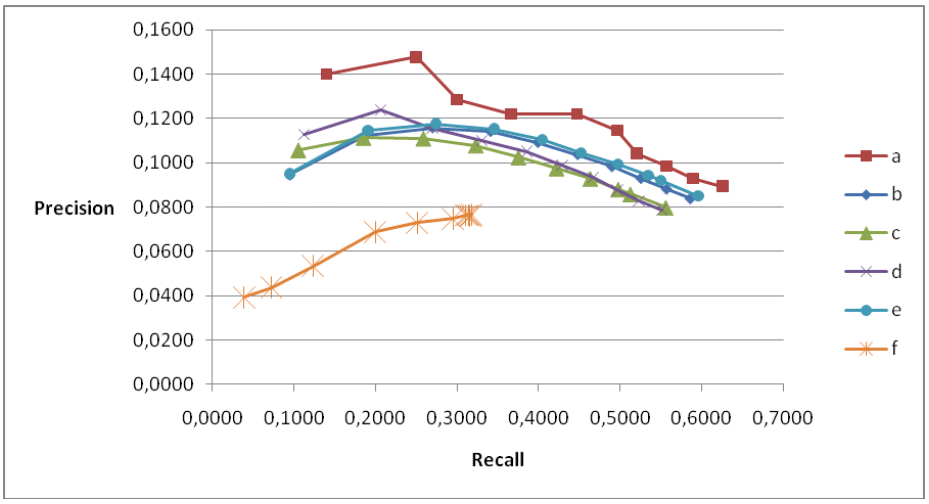

Fig. 2. Precision/Recall curves for a. Latent Baskets – Gibbs Sampler, b. Latent Baskets Thesaurus, c. Latent Baskets with Co-occurrence Boosting, d. Latent Users Thesaurus, e. Latent Baskets combined with Latent Users and f. FP-Growth Association Rules.

The curves in Fig. 2 indicate that all topic model based recommender techniques evaluated in this work significantly outperform association rules mining. It is notable that: (a) the Gibbs Sampling-enabled topic model recommender successfully retrieves more than 60% of the bought items on average. However, Gibbs Sampling, which involves statistical inference, is inefficient in terms of storage size and computational power. Alternative thesaurus based approaches (b), (c), (d) and (e) provide satisfactory results compared to (a) and therefore can be used in practice.

Fig.2 also exhibits the differences in the results of the thesaurus based techniques. Both approaches for analysing transactions, either considering only transactions (Latent Baskets) and taking user history into consideration (Latent Users), generate topic models that can successfully predict user behaviour. Specifically, Latent User topic models (d) provide more precise recommendations earlier than Latent Baskets (b). This effect, however, is reversed as the recommendation size gets bigger (see Table 6). The performance of the Latent Basket topic models is slightly improved by increasing the score of items that co-exist in multiple topics (c). Nevertheless, as the number of recommendations increases it leads to worse results. Our integrated method of using a linear combination of Latent Users and Baskets (e) provides an effective recommendation system for any number of recommendations. Finally, with regards to the FP-Growth generated rules (f), we notice that association rules

precision and recall remains stable when the recommendation size rises, as the number of applicable rules cannot provide the recommendations required.

6 Conclusions and Future Work

In this paper, we apply latent topic models in purchase data in order to provide a model for consumer behaviour as well as to recommend products to consumers. Our research indicates that latent topic analysis is an appropriate and effective method for analysing purchase data, one that outperforms association rule mining, a technique typically used in market basket analysis. Latent topic analysis does not only provide insight into the consumer preferences but also effectively support an item recommendation system. Specifically, the capability to discover latent baskets and latent users lays the foundation for understanding user tastes and their relations to items. Moreover, latent topic models can be effective in recommending items to users even when applied to large data sets and large itemsets.

In the future we will consider two different lines of further research. First, variations of the Latent Dirichlet Allocation can be used in the process of modelling customer preferences. Specifically Hierarchical LDA [4] can be combined with the current product hierarchy, as well as methods used in social media, such as Labeled LDA [31]. Second, it is interesting to see how the topic models can be adapted to accommodate a recommender for a web based marketplace, where consumers can also explicitly rank the available products.

References

1. Adomavicius, G., Tuzhilin, A.: Toward the next generation of recommender systems: a survey of the state-ofthe-art and possible extensions, Knowledge and Data Engineering, IEEE Transactions, vol. 17, 734—749 (2005).
2. Agrawal, R., Imielinski, T., Swami, A.: Mining association rules between sets of items in large databases, ACM SIGMOD Record, vol. 22, 207--216 (1993)
3. Bíró I, Szabó J, Benczúr AA. Latent dirichlet allocation in web spam filtering. Proceedings of the 4th international workshop on Adversarial information retrieval on the web, Beijing, China: ACM,29-32 (2008)
4. Blei, D., Griffiths, T., Jordan, M., & Tenenbaum, J.: Hierarchical topic models and the nested Chinese restaurant process. NIPS (2004)
5. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation, In Journal of Machine Learning Research, vol. 3, 993--1022 (2003)
6. Bose, I., Chen, X., Quantitative models for direct marketing: A review from systems perspective, European Journal of Operational Research, Vol. 195, Issue 1, 1—16 (2009)
7. Burke, R.: Hybrid recommender systems: Survey and experiments, User Modeling and User-Adapted Interaction, vol. 12, 331--370 (2002)
8. Calders, T. Goethals, B.: Non-derivable itemset mining, Data Mining and Knowledge Discovery, vol. 14, 171—206 (2007)
9. Cavique, L.: A scalable algorithm for the market basket analysis, Journal of Retailing and Consumer Services 14, 400—407 (2007)

10. Chen, Y.-L., Tang, K., Hu, Y.-H.: Market basket analysis in a multiple store environment. Decision Support Systems 40 (2), 339--354 (2005)
11. Cho, Y.: A personalized recommender system based on web usage mining and decision tree induction, Expert Systems with Applications, vol. 23, 329—342 (2002)
12. Data Mining: Practical Machine Learning Tools and Techniques, Ian H. Witten, Eibe Frank , Morgan Kaufmann , June 2005, p. 235
13. Gregg, D., Walczak, S.: Auction Advisor: an agent-based online-auction decision support system, Decision Support Systems, vol. 41, 449—471 (2006)
14. Han, I., Pei, J., Yin, J., Mao, R.: Mining Frequent Patterns without Candidate Generation: A Frequent-Pattern Tree Approach, Data Mining and Knowledge Discovery, vol. 8, 53--87 (2004)
15. Hofmann, T.: Probabilistic latent semantic indexing," In Proc. 22nd Annual Int'l. ACM SIGIR Conf. on Research and Development in Information Retrieval. Kitagawa, G, vol. 5, 1—25 (1999)
16. Huang, Z., Zeng, D., Chen, H.: A comparison of collaborative-filtering recommendation algorithms for e-commerce, IEEE Intelligent Systems, vol. 22, 68--78 (2007)
17. Huang, Z., Zhung, W., Chen, H.: A graph model for E-commerce recommender systems, Journal of the American Society for Information Science and Technology, vol. 55, 259--274 (2004).
18. Iwata, T., Watanabe, S., Yamada, T., Ueda, N.: Topic Tracking Model for Analyzing Consumer Purchase Behavior, International Joint Conference on Artificial Intelligence, 2---4 (2009)
19. Furnkranz J. , Hullermeier E., "Preference Learning: An Introduction," Preference Learning, To appear
20. Krestel R, Fankhauser P, Nejdl W. Latent Dirichlet Allocation for Tag Recommendation. Proceedings of the third ACM conference on Recommender systems, 61–68 (2009)
21. Lee, J., Jun, C., Kim, S.: Classification-based collaborative filtering using market basket data, Expert Systems with Applications, vol. 29, 700—704 (2005)
22. Lee, W.: Towards agent-based decision making in the electronic marketplace: interactive recommendation and automated negotiation, Expert Systems with Applications, vol. 27, 665—679 (2004)
23. Linden, G., Smith, B., York, J.: Amazon.com Recommendations: Item-to-Item Collaborative Filtering, IEEE Internet Computing, Jan./Feb (2003)
24. Lukins SK, Kraft NA, Etzkorn LH. Source Code Retrieval for Bug Localization Using Latent Dirichlet Allocation, IEEE Computer Society,155-164 (2008)
25. McCallum, A.K. "MALLET: A Machine Learning for Language Toolkit." http://mallet.cs.umass.edu. 2002.
26. Mild, A., Reutterer, T.: An improved collaborative filtering approach for predicting cross-category purchases based on binary market basket data, Journal of Retailing and Consumer Services, vol. 10, 123—133 (2003)
27. Ohsawa, Y., Yada, K.: Data Mining for Design and Marketing, CRC Press (2009)
28. Olson, D., Delen, D.: Advanced data mining techniques, Springer Verlag (2008)
29. Park, L. Ramamohanarao, K.: Efficient storage and retrieval of probabilistic latent semantic information for information retrieval, The VLDB Journal, vol. 18, 141—155 (2009)
30. R. Agrawal, T. Imielinski, and A. Swami, "Mining association rules between sets of items in large databases," *ACM SIGMOD Record*, vol. 22, 207–216 (1993)
31. Ramage, D., Hall, D., Nallapati, R., Manning. C.D.: Labeled lda: A supervised topic model for credit attribution in multi-labeled corpora. In Proceddings of the 2009 Conference on Empirical Methods in Natural Language Processing (2009)
32. RapidMiner http://rapid-i.com/

33. Sarwar, B., Karypis, G., Konstan, J., Riedl, J.: Analysis of recommendation algorithms for e-commerce, Proceedings of the 2nd ACM conference on Electronic commerce, ACM, 167 (2000)
34. Savasere, A., Omiccinski, E., Navathe, S.: An efficient algorithm for mining association rules in large databases, Proceedings of the International Conference on Very Large Data Bases, Citeseer, 432--444 (1995)
35. Schafer, B.J., Konstan, J.A., Riedi, J.: Recommender systems in e-commerce, In ACM Conference on Electronic Commerce, 158—166 (1999)
36. Song, I., Chintagunta, P.K.: Measuring cross-category price effects with aggregate store data. Management Science 52 (10), 1594—1609 (2006)