Two-Way Grouping by One-Way Topic Models

Eerika Savia, Kai Puolamäki, and Samuel Kaski

Helsinki Institute for Information Technology HIIT
Department of Information and Computer Science
Helsinki University of Technology
P.O. Box 5400, FI-02015 TKK, Finland
{forename.surname}@tkk.fi

Abstract. We tackle the problem of new users or documents in collaborative filtering. Generalization over users by grouping them into user groups is beneficial when a rating is to be predicted for a relatively new document having only few observed ratings. The same applies for documents in the case of new users. We have shown earlier that if there are both new users and new documents, two-way generalization becomes necessary, and introduced a probabilistic Two-Way Model for the task. The task of finding a two-way grouping is a non-trivial combinatorial problem, which makes it computationally difficult. We suggest approximating the Two-Way Model with two URP models; one that groups users and one that groups documents. Their two predictions are combined using a product of experts model. This combination of two one-way models achieves even better prediction performance than the original Two-Way Model.

1 Introduction

This paper considers models for the task of predicting relevance values for user–item pairs based on a set of observed ratings of users for the items. In particular, we concentrate on the task of predicting relevance when very few ratings are known for each user or item.[1]

In so-called collaborative filtering methods the predictions are based on the opinions of similar-minded users. Collaborative filtering is needed when the task is to make personalized predictions but there is not enough data available for each user individually. The early collaborative filtering methods were memory-based (see, e.g., [1,2]). Model-based approaches are justified by the poor scaling of the memory-based techniques. Recent work includes probabilistic and information-theoretic models, see for instance [3,4,5,6].

A family of models most related to our work are the latent topic models, which have been successfully used in document modeling but also in collaborative filtering [7,8,9,10,11,12,13,14,15,16,17]. The closest related models include probabilistic Latent Semantic Analysis (pLSA; [3]), Latent Dirichlet Allocation

---

[1] The models we discuss are generally applicable, but since our prototype application area has been information retrieval we will refer to the items as documents.

(LDA; [18,19]), and User Rating Profile model (URP; [20]), which all assume a one-way grouping. In addition, there is a two-way grouping model, called Flexible Mixture Model (FMM; [21]). We have discussed the main differences between our Two-Way Model and these related models in [22].

1.1   Cold-Start Problem

Since a collaborative filtering system has to rely on the past experiences of the users, it will have problems when assessing new documents that have not yet been seen by most of the users. Making the collaborative filtering scheme item-based, that is, grouping items or documents instead of users, would in turn imply the problem where new users that have only few ratings will get poor predictions. This problem of unseen or almost unseen users and documents is generally referred to as the *cold-start problem* in recommender system literature, see for instance [23]. The Two-Way Model was proposed to tackle this problem of either new users or new documents [22,24].

1.2   Approximating Two-Way Model with Two One-Way Models

It has been shown for hard biclustering of binary data matrices, that clustering the marginals independently to produce a check-board-like biclustering is guaranteed to achieve fairly good results compared to the NP-hard optimal solution. An approximation ratio for the crossing of two one-way clusterings has been proven [25,26]. Inspired by this theoretical guarantee, we suggest approximating the Two-Way Model with two User Rating Profile models (URP, [20]); one that groups users and one that groups documents. The combination of the two Gibbs-sampled probabilistic predictions is made using a product of experts model [27].

We have followed the experimental setups of our earlier study [22] in order to be able to compare the results in a straightforward manner. We briefly describe the experimental scenarios, the performance measures and the baseline models in Sect. 3. In Sect. 4.1 we demonstrate with clearly clustered toy data how the product of two URP models improves the relevance predictions of the corresponding one-way models. Finally, in Sect. 4.2 we show in a real-world case study from our earlier paper that the proposed method works as expected also in practice.

We expected the proposed method to have the advantage of giving better predictions than the individual one-way models with the computational complexity of the one-way model. The one-way grouping models are faster and more reliable in their convergence than the Two-Way Model, basically because of the difference in the intrinsic complexity of the tasks they are solving.

2   Method

Originally, User Rating Profile model was suggested to be estimated by variational approximation (variational URP, [20]), but we have introduced also Gibbs-sampled variants of the model in [22,24] (Gibbs URP and Gibbs URP-GEN).

The difference between a one-way model and the Two-Way Model is whether to cluster only users (documents) or to cluster both users and documents. Another difference between URP and the Two-Way Model is whether the users and documents are assumed to be generated by the model or treated as covariates of the model. In our earlier study [22] it was found that unless the data marginals are especially misleading about the full data, it is always useful to design the model to be fully generative, in contrast to seeing users and documents as given covariates of the model. Therefore, we have only included the generative variants of Gibbs URP models is this study (Gibbs URP-GEN).

2.1   One-Way Grouping Models

In Fig. 1 we show graphical representations of the generative Gibbs URP model introduced in [22] (User Gibbs URP-GEN), and the corresponding document-grouping variant (Doc Gibbs URP-GEN). They are the one-way grouping models used as the basis of our suggested method. Our main notations are summarized in Table 1.

2.2   Two-Way Grouping Model

In Fig. 2 we show a graphical representation of the Two-Way Model that our suggested method approximates. The Two-Way Model generalizes the generative user-grouping URP by grouping both users and documents. It has been shown to predict relevance more accurately than one-way models when the target consists of both new documents and new users. The reason is that generalization over documents becomes beneficial for new documents and at the same time generalization over users is needed for new users. Finally, Table 2 summarizes the differences between the models.

2.3   Approximation of Two-Way Model by Product of Experts

We propose a model where we estimate predictive Bernoulli distributions separately with user-based URP and document-based URP and combine their results with a product of experts model [27]. To be exact, we took the product of the Bernoulli relevance probabilities given by the user-based URP ($P_U(r = 1|u, d)$) and the document-based URP ($P_D(r = 1|u, d)$) and normalized the product distributions, as follows:

$$P_{PoE}(r = 1|u, d) = \frac{P_U(r = 1|u, d) \, P_D(r = 1|u, d)}{\sum_{r=0,1} P_U(r|u, d) \, P_D(r|u, d)} \quad . \tag{1}$$

2.4   Baseline Models

We compared our results to two simple baseline models. These models mainly serve as an estimate of the lower bound of performance by making an assumption that the data comes from one cluster only. The *Document Frequency Model* does not take into account differences between users or user groups at all. It simply

Table 1. Notation

SYMBOL	DESCRIPTION
u	user index
d	document index
r	binary relevance (relevant = 1, irrelevant = 0)
u^*	user group index (attitude in URP)
d^*	document cluster index
N_U	number of users
N_D	number of documents
N	number of triplets (u, d, r)
K_U	number of user groups
K_D	number of document clusters
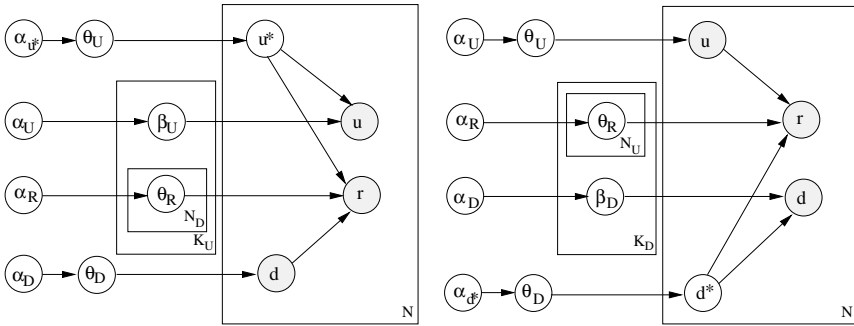
(a) User Gibbs URP-GEN groups only users and assumes that the relevance depends solely on the user group and the document.

(b) Doc Gibbs URP-GEN groups only documents and assumes that the relevance depends solely on the document cluster and the user.

Fig. 1. Graphical model representations of the generative Gibbs URP models with user grouping (User Gibbs URP-GEN) and with document grouping (Doc Gibbs URP-GEN). The grey circles indicate observed values. The boxes are "plates" representing replicates; the value in a corner of each plate is the number of replicates. The rightmost plate represents the repeated choice of N (user, document, rating) triplets. The plate labeled with K_U (or K_D) represents the different user groups (or document clusters), and $\boldsymbol{\beta}_U$ (or $\boldsymbol{\beta}_D$) denotes the vector of multinomial parameters for each user group (or document cluster). The plate labeled with N_D (or N_U) represents the documents (or users). In the intersection of these plates there is a Bernoulli-model for each of the $K_U \times N_D$ (or $K_D \times N_U$) combinations of user group and document (or document cluster and user). Since α_D and θ_D (or α_U and θ_U) are conditionally independent of all other parameters given document d (or user u), they have no effect on the predictions of relevance $P(r \mid u, d)$ in these models. They only describe how documents d (or users u) are assumed to be generated. A table listing distributions of all the random variables can be found in the Appendix.
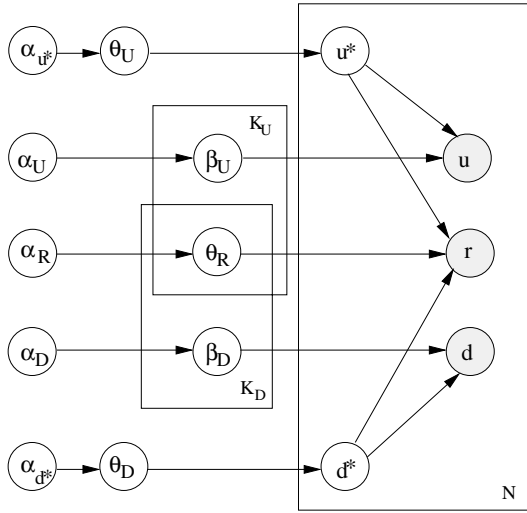
Fig. 2. Graphical model representation of the Two-Way Model, which groups both users and documents and assumes that the relevance depends only on the user group and the document cluster instead of individual users/documents. The rightmost plate represents the repeated choice of N (user, document, rating) triplets. The plate labeled with K_U represents the different user groups, and $\boldsymbol{\beta}_U$ denotes the vector of multinomial parameters for each user group. The plate labeled with K_D represents the different document clusters, and $\boldsymbol{\beta}_D$ denotes the vector of multinomial parameters for each document cluster. In the intersection of these plates there is a Bernoulli-model for each of the $K_U \times K_D$ combinations of user group and document cluster. A table listing distributions of all the random variables can be found in the Appendix.

Table 2. Summary of the models (**u**=user, **d**=document). The column "**Gibbs**" indicates which of the models are estimates by Gibbs sampling, in contrast to variational approximation. Prefix "2-way" stands for combination of two one-way models.

Model Abbreviation	Generates u,d	Gibbs	Groups u	Groups d
Two-Way Model	●	●	●	●
2-way Gibbs URP-GEN	●	●	●	●
2-way Gibbs URP	–	●	●	●
2-way Variational URP	–	–	●	●
1-way User Gibbs URP-GEN	●	●	●	–
1-way User Gibbs URP	–	●	●	–
1-way User Var URP	–	–	●	–
1-way Doc Gibbs URP-GEN	●	●	–	●
1-way Doc Gibbs URP	–	●	–	●
1-way Doc Var URP	–	–	–	●

models the probability of a document being relevant as the frequency of $r = 1$ in the training data for the document:

$$P(r = 1 \mid d) = \frac{\sum_u \#(u, d, r = 1)}{\sum_{u,r} \#(u, d, r)} \quad . \tag{2}$$

The *User Frequency Model*, on the other hand, does not take into account differences between documents or document groups. It is the analogue of Document Frequency Model, where the roles of users and documents have been interchanged.

3   Experiments

3.1   Experimental Scenarios

In this section we describe the different types of experimental scenarios that were studied with both data sets. The training and test sets were taken from the earlier study [22]. The scenarios have various levels of difficulty for models that group only users, only documents, or that group both.

- **Only "New" Documents.** This scenario had been constructed to correspond to prediction of relevances for new documents in information retrieval. It had been taken care that each of the randomly selected test documents had only 3 ratings in the training data. The rest of the ratings for these documents had been left to the test set. For the rest of the documents, all the ratings were included in the training set. Hence, the models were able to use "older" documents (for which users' opinions are already known) for training the user groups and document clusters. This scenario favors models that cluster documents.
- **Only "New" Users.** The experimental setting for new users had been constructed in exactly the same way as the setting for new documents but with the roles of users and documents reversed. This scenario favors models that cluster users.
- **Either User or Document is "New".** In an even more general scenario either the users or the documents can be "new." In this setting the test set consisted of user-document pairs where either the user is "new" and the document is "old" or vice versa. This scenario brings out the need for two-way generalization.
- **Both User and Document are "New".** In this setting all the users and documents appearing in the test set were "new," having only 3 ratings in the training set. This case is similar to the previous setting but much harder, even for the two-way grouping models.

3.2   Measures of Performance

For all the models, we used log-likelihood of the test data set as a measure of performance, written in the form of perplexity,

$$\text{perplexity} = e^{-\frac{\mathcal{L}}{N}} \text{ , where } \mathcal{L} = \sum_{i=1}^{N} \log P(r_i \mid u_i, d_i, \mathcal{D}) \quad . \tag{3}$$

Here $\mathcal{D}$ denotes the training set data, and N is the size of the test set. Gibbs sampling gives an estimate for the table of relevance probabilities over all (u, d) pairs, $P(r \mid u, d, \mathcal{D})$, from which the likelihood of each test pair (u_i, d_i) can be estimated as $P(r_i \mid u_i, d_i, \mathcal{D})$.[2]

We further computed the accuracy, that is, the fraction of the triplets in the test data set for which the prediction was correct. We took the predicted relevance to be $\arg\max_{r \in \{0,1\}} P(r \mid u, d, \mathcal{D})$, where $P(r \mid u, d, \mathcal{D})$ is the probability of relevance given by the model. Statistical significance was tested with the Wilcoxon signed rank test.

3.3   Demonstration with Artificial Data

The artificial data sets were taken from the earlier study [22]. The experimental setting is described in detail in the technical report [28]. In brief, the data was designed such that it contained bicluster structure with $K_U = K_D = 3$. There were 10 artificial data sets of size 18,000, that all followed the same bicluster structure.

All the models were trained with the known true numbers of clusters. For each of the 10 data sets the models were trained with a training set and tested with a separate test set, and the final result was the mean of the 10 test set perplexities.

The generative Gibbs URP models were combined as a product of experts model. According to our earlier studies, the variational URP generally seems to produce extreme predictions, near either 0 or 1. Therefore, the variational URP models (User Var URP and Doc Var URP) were combined as a hard biclustering model, as follows. The MAP estimates for cluster belongings from the distributions of the one-way variational URP models were used to divide all the users and documents into bins to produce a hard check-board-like biclustering. In each bicluster the $P(r = 1|u, d)$ was set to the mean of the training data points that lay in the bicluster.

3.4   Experiments with Parliament Data

We selected the cluster numbers using a validation set described in [28]. The validated cluster numbers (Two-Way Model $K_U = 4$ and $K_D = 2$, User Gibbs URP-GEN $K_U = 2$, Doc Gibbs URP-GEN $K_D = 2$) were used in all experimental scenarios. The choices from which the cluster numbers were selected were $K_U \in \{1, 2, 3, 4, 5, 10, 20, 50\}$ for the user groups and $K_D \in \{1, 2, 3, 4, 5, 10, 20\}$ for the document clusters.

4   Results

4.1   Results of Experiments with Artificial Data

The results of the experiment with artificial data are shown in Table 3. The proposed product of two generative Gibbs URP models outperformed even the

---

[2] Theoretically, perplexity can grow without a limit if the model predicts zero probability for some element in the test data set, so in practice, we clipped the probabilities to the range $[e^{-10}, 1]$.

Two-Way Model in all the scenarios, being the best in all but the "both new" case, where the hard clustering of MAP estimates of variational URP models was the best. The hard biclustering model worked very well for the variational URP (See Table 3), in contrast to the product of experts -combination, which did not perform well for variational URP[3]. The prediction accuracy of the best model varied between 83–84%, while the prediction accuracy of the best baseline model varied between 50–52%. The full results with all the accuracy values can be found in the technical report [28].

Table 3. Perplexity of the various models in experiments with artificial data. In each column, the best model (underlined) differs statistically significantly from the second-best one (P-value ≤ 0.01). Small perplexity is better; 2.0 corresponds to binary random guessing and 1.0 to perfect prediction.

Method	New Doc	New User	Either New	Both New
Two-Way Model	1.52	1.54	1.53	1.70
2-way Gibbs URP-GEN	<u>1.46</u>	<u>1.47</u>	<u>1.45</u>	1.70
2-way Var URP	1.55	1.57	1.54	<u>1.52</u>
User Gibbs URP-GEN	1.68	1.57	1.62	1.83
User Var URP	7.03	2.07	3.45	9.27
Doc Gibbs URP-GEN	1.56	1.69	1.62	1.81
Doc Var URP	1.86	5.99	3.08	6.90
User Freq.	2.02	5.65	3.25	4.99
Document Freq.	5.29	2.01	3.21	5.92

4.2   Results of Experiments with Parliament Data

The product of two generative Gibbs URP models outperformed even the Two-Way Model in all the scenarios, being the best in all cases (see Table 4). The prediction accuracy of the best model varied between 93–97%, while the prediction accuracy of the best baseline model varied between 64–71%. The full results with all the accuracy values can be found in the technical report [28].

5   Discussion

We have tackled the problem of new users or documents in collaborative filtering. We have shown in our previous work that if there are both new users and new documents, two-way generalization becomes necessary, and introduced a probabilistic Two-Way Model for the task in [22].

In this paper we suggest an approximation for the Two-Way Model with two User Rating Profile models — one that groups users and one that groups

---

[3] We only show the performance of Variational URP for the artificial data since our implementation is too inefficient for larger data sets.

Table 4. Parliament Data. Comparison between the models by perplexity over the test set. In each column, the best model (underlined) differs statistically significantly from the second-best one (P-value $\leq$ 0.01). Small perplexity is better; 2.0 corresponds to binary random guessing and 1.0 to perfect prediction.

Method	New Doc	New User	Either New	Both New
Two-Way Model	1.37	1.40	1.38	1.62
2-way Gibbs URP-GEN	<u>1.19</u>	<u>1.22</u>	<u>1.20</u>	<u>1.45</u>
User Gibbs URP-GEN	1.47	1.34	1.41	1.64
Doc Gibbs URP-GEN	**1.34**	1.54	1.43	1.68
User Freq.	2.00	5.68	3.32	4.78
Document Freq.	5.36	1.76	3.12	5.85

documents — which are combined as a product of experts (PoE). We show with two data sets from the earlier study [22], that the PoE model achieves the performance level of the more principled Two-Way Model and even outperforms it.

The task of finding such a two-way grouping that best predicts the relevance is a difficult combinatorial problem, which makes convergence of the sampling hard to achieve. This work was motivated by the finding that hard biclustering of binary data can be approximated using two one-way clusterings with a proven approximation ratio.

The main advantage of the proposed method, compared to earlier works, is the ability to make at least as good predictions as the Two-Way Model but with the computational complexity of the one-way model. We assume that the reason why the product of experts combination outperformed the Two-Way Model lies in the less reliable and slower convergence of the Two-Way Model compared to the one-way grouping models. This is basically due to the difference in the intrinsic complexity of the tasks they are solving.

Acknowledgments

References

1. Konstan, J., Miller, B., Maltz, D., Herlocker, J.: GroupLens: Applying collaborative filtering to usenet news. Communications of the ACM 40(3), 77–87 (1997)
2. Shardanand, U., Maes, P.: Social information filtering: Algorithms for automating 'word of mouth'. In: Proceedings of the ACM CHI 1995 Human Factors in Computing Systems Conference, pp. 210–217 (1995)

3. Hofmann, T.: Latent semantic models for collaborative filtering. ACM Trans. Inf. Syst. 22(1), 89–115 (2004)

4. Jin, R., Si, L.: A Bayesian approach towards active learning for collaborative filtering. In: Proceedings of the Twentieth Conference on Uncertainty in Artificial Intelligence, UAI 2004, pp. 278–285. AUAI Press (2004)

5. Wettig, H., Lahtinen, J., Lepola, T., Myllymäki, P., Tirri, H.: Bayesian analysis of online newspaper log data. In: Proc. of the 2003 Symposium on Applications and the Internet Workshops (SAINT 2003), pp. 282–287. IEEE Computer Society, Los Alamitos (2003)

6. Zitnick, C., Kanade, T.: Maximum entropy for collaborative filtering. In: Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence, UAI 2004, pp. 636–643. AUAI Press (2004)

7. Blei, D.M., Jordan, M.I.: Modeling annotated data. In: Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 127–134. ACM Press, New York (2003)

8. Buntine, W., Jakulin, A.: Discrete component analysis. In: Saunders, C., Grobelnik, M., Gunn, S., Shawe-Taylor, J. (eds.) SLSFS 2005. LNCS, vol. 3940, pp. 1–33. Springer, Heidelberg (2006)

9. Erosheva, E., Fienberg, S., Lafferty, J.: Mixed membership models of scientific publications. Proc. of the National Academy of Sciences 101, 5220–5227 (2004)

10. Keller, M., Bengio, S.: Theme topic mixture model: A graphical model for document representation. In: PASCAL Workshop on Text Mining and Understanding (2004)

11. Marlin, B., Zemel, R.S.: The multiple multiplicative factor model for collaborative filtering. In: ICML 2004: Proceedings of the 21th International Conference on Machine Learning, p. 73. ACM Press, New York (2004)

12. McCallum, A., Corrada-Emmanuel, A., Wang, X.: The author-recipient-topic model for topic and role discovery in social networks: Experiments with Enron and Academic Email. Technical report, University of Massachusetts (2004)

13. Popescul, A., Ungar, L., Pennock, D., Lawrence, S.: Probabilistic models for unified collaborative and content-based recommendation in sparse-data environments. In: Proceedings of UAI 2001, pp. 437–444. Morgan Kaufmann, San Francisco (2001)

14. Pritchard, J.K., Stephens, M., Donnelly, P.: Inference of population structure using multilocus genotype data. Genetics 155, 945–959 (2000)

15. Rosen-Zvi, M., Griffiths, T., Steyvers, M., Smyth, P.: The author-topic model for authors and documents. In: Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence, UAI 2004, pp. 487–494. AUAI Press (2004)

16. Yu, K., Yu, S., Tresp, V.: Dirichlet enhanced latent semantic analysis. In: Cowell, R.G., Ghahramani, Z. (eds.) Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics, AISTATS 2005, pp. 437–444. Society for Artificial Intelligence and Statistics (2005)

17. Yu, S., Yu, K., Tresp, V., Kriegel, H.-P.: A probabilistic clustering-projection model for discrete data. In: Jorge, A.M., Torgo, L., Brazdil, P.B., Camacho, R., Gama, J. (eds.) PKDD 2005. LNCS (LNAI), vol. 3721, pp. 417–428. Springer, Heidelberg (2005)

18. Blei, D., Ng, A.Y., Jordan, M.I.: Latent Dirichlet allocation. Journal of Machine Learning Research 3, 993–1022 (2003)

19. Buntine, W.: Variational extensions to EM and multinomial PCA. In: Elomaa, T., Mannila, H., Toivonen, H. (eds.) ECML 2002. LNCS (LNAI), vol. 2430, pp. 23–34. Springer, Heidelberg (2002)

20. Marlin, B.: Modeling user rating profiles for collaborative filtering. In: Advances in Neural Information Processing Systems 16, pp. 627–634. MIT Press, Cambridge (2004)
21. Si, L., Jin, R.: Flexible mixture model for collaborative filtering. In: Fawcett, T., Mishra, N. (eds.) Proceedings of the Twentieth International Conference on Machine Learning, ICML 2003, pp. 704–711. AAAI Press, Menlo Park (2003)
22. Savia, E., Puolamäki, K., Kaski, S.: Latent grouping models for user preference prediction. Machine Learning 74(1), 75–109 (2009)
23. Lam, X.N., Vu, T., Le, T.D., Duong, A.D.: Addressing cold-start problem in recommendation systems. In: ICUIMC 2008: Proceedings of the 2nd international conference on Ubiquitous information management and communication, pp. 208–211. ACM, New York (2008)
24. Savia, E., Puolamäki, K., Sinkkonen, J., Kaski, S.: Two-way latent grouping model for user preference prediction. In: Bacchus, F., Jaakkola, T. (eds.) Uncertainty in Artificial Intelligence 21, pp. 518–525. AUAI Press, Corvallis (2005)
25. Puolamäki, K., Hanhijärvi, S., Garriga, G.C.: An approximation ratio for biclustering. Information Processing Letters 108, 45–49 (2008)
26. Anagnostopoulos, A., Dasgupta, A., Kumar, R.: Approximation algorithms for co-clustering. In: Proceedings of the Twenty-Seventh ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems, pp. 201–210. ACM, New York (2008)
27. Hinton, G.E.: Training Products of Experts by Minimizing Contrastive Divergence. Neural Computation 14(8), 1771–1800 (2002)
28. Savia, E., Puolamäki, K., Kaski, S.: On two-way grouping by one-way topic models. Technical Report TKK-ICS-R15, Helsinki University of Technology, Department of Information and Computer Science, Espoo, Finland (May 2009)

A   Distributions in the Models

Table 5. Summary of the distributions in User Gibbs URP-GEN

SYMBOL	DESCRIPTION
$\boldsymbol{\beta}_U(u^*)$	Vector of multinomial parameters defining the probabilities of certain user group u^* to contain each user
$\boldsymbol{\theta}_U$	Multinomial probabilities of user groups u^* to occur
$\boldsymbol{\theta}_D$	Multinomial probabilities of documents d to occur (needed only for the generative process)
$\boldsymbol{\theta}_R(u^*, d)$	Vector of Bernoulli parameters defining the probabilities of certain user group u^* to consider document d relevant or irrelevant
$\boldsymbol{\alpha}_U$	Dirichlet prior parameters for all $\boldsymbol{\beta}_U$
$\boldsymbol{\alpha}_{u^*}$	Dirichlet prior parameters for $\boldsymbol{\theta}_U$
$\boldsymbol{\alpha}_D$	Dirichlet prior parameters for $\boldsymbol{\theta}_D$ (needed only for the generative process)
$\boldsymbol{\alpha}_R$	Dirichlet prior parameters for all $\boldsymbol{\theta}_R$

Table 6. Summary of the distributions in the Two-Way Model

SYMBOL	DESCRIPTION
$\boldsymbol{\theta}_U$	Multinomial probabilities of user groups u^* to occur
$\boldsymbol{\beta}_U(u^*)$	Vector of multinomial parameters defining the probabilities of certain user group u^* to contain each user
$\boldsymbol{\theta}_D$	Multinomial probabilities of document clusters d^* to occur
$\boldsymbol{\beta}_D(d^*)$	Vector of multinomial parameters defining the probabilities of certain document cluster d^* to contain each document
$\boldsymbol{\theta}_R(u^*, d^*)$	Vector of Bernoulli parameters defining the probabilities of certain user group u^* to consider document cluster d^* relevant or irrelevant
$\boldsymbol{\alpha}_U$	Dirichlet prior parameters for all $\boldsymbol{\beta}_U$
$\boldsymbol{\alpha}_{u^*}$	Dirichlet prior parameters for $\boldsymbol{\theta}_U$
$\boldsymbol{\alpha}_D$	Dirichlet prior parameters for all $\boldsymbol{\beta}_D$
$\boldsymbol{\alpha}_{d^*}$	Dirichlet prior parameters for $\boldsymbol{\theta}_D$
$\boldsymbol{\alpha}_R$	Dirichlet prior parameters for all $\boldsymbol{\theta}_R$