

INTRODUCING A STAR TOPOLOGY INTO LATENT CLASS MODELS FOR COLLABORATIVE FILTERING

Gabriela Polčicová

Faculty of Informatics and Information Technologies

*Slovak University of Technology, **Ilkovičova** 3, 81219 Bratislava, Slovakia*

polcicova@fiit.stuba.sk

Peter Tiño

School of Computer Science

University of Birmingham, Birmingham, B15 2TT, UK

P.Tino@cs.bham.ac.uk

Abstract Latent class models (LCM) represent the high dimensional data in a smaller dimensional space in terms of latent variables. They are able to automatically discover the patterns from the data. We present a topographic version of two LCMs for collaborative filtering and apply the models to a large collection of user ratings for films. Latent classes are topologically organized on a “star-like” structure. This makes orientation in rating patterns captured by latent classes easier and more systematic. The variation in film rating patterns is modelled by multinomial and binomial distributions with varying independence assumptions.

Keywords: collaborative filtering, latent variable models, visualization

Introduction

When deciding which book to read, which film to watch, or which website to visit, people often rely on advice given by other people [3]. This is possible only inside small communities, where people know each other. In many situations we would like to automate that process. The method addressing the problem of making recommendations is collaborative filtering (CF) that leverages the existing preferences (ratings/profiles) of users in large communities.

One approach to CF uses probabilistic modeling (e.g. LCM) to infer new recommendations. The main advantage of this approach is that it is able to automatically discover preference patterns in user profile data.

Besides producing new recommendations, the CF system can be used to understand the principal taste trends in the rating data by careful analysis of preference patterns extracted by LCM. Unfortunately, detailed examination of all latent classes is a very tiring and time consuming process. In this paper we suggest a way of addressing this problem: the latent space is endowed with a topological organization that enables us to visualize the common interest patterns in an easily accessible way. We study latent class models for user ratings with star topology.

The paper is organized as follows. In section 1 we describe LCMs for user ratings and in Section 2 we endow them with star topology. Section 3 is devoted to Expectation-Maximization (EM) algorithm for training the models. The experiments are described in section 4. The paper is concluded in section 5.

1. Latent Class Models for User Ratings

In this section we briefly describe latent class approach to modeling user ratings introduced in [3, 4]. We work with three sets: the set \mathcal{U} of users, the set of films (items), \mathcal{Y} , and the set \mathcal{V} of rating values that are used by users to evaluate films.

We would like to predict the rating $v_{u,y} \in \mathcal{V}$ given by a user $u \in \mathcal{U}$ to a film $y \in \mathcal{Y}$. With each triplet $(u, y, v_{u,y})$, either observed or just hypothetical, we associate a latent variable (class) $z_{u,y} \in \mathcal{Z} = \{1, 2, \dots, K\}$ that “explains” why the user u rates the film y by $v_{u,y}$.

The latent variables $z \in \mathcal{Z}$ index “abstract” classes of users in two types of models:

- **Type I** – given a film y , all users from class z tend to adopt the same rating pattern expressed through $P(v|y, z)$ over evaluations from \mathcal{V} . Given a user u and a film y , the probability of vote v is modeled as $P(v|y, u) = \sum_{z \in \mathcal{Z}} P(v|y, z)P(z|u)$, where $P(z|u)$ is the probability that the user u “participates” in class z .
- **Type II** – all users from class z tend to adopt the same preferences over the [rating, film] pairs (v, y) . Given a user u , the probability of a pair (v, y) is modeled as $P(v, y|u) = \sum_{z \in \mathcal{Z}} P(v, y|z)P(z|u)$.

Classes z express “common interest patterns” among the users found in the ratings [3] and $P(z|u)$ then represents to what extent the user u participates in the common interest pattern z .

Given a set of observation triplets, free parameters of the model, $P(\mathbf{z}|\mathbf{u})$ and $P(\mathbf{v}|\mathbf{y}, \mathbf{z})$, are determined by an EM procedure [3].

2. Introducing a Topology into Latent Class Models

We endow the latent classes with a topographic organization. In [5] we presented topographic organization of latent classes on a grid. In this paper we introduce a less tight star topology that organizes latent classes such that similar classes lie close to each other (on the same branch).

Topology is introduced into the latent space via the channel noise methodology [2]: for latent classes \mathbf{z}_1 and \mathbf{z}_2 lying close to each other, the probability $P(\mathbf{z}_2|\mathbf{z}_1) = P(\mathbf{z}_1|\mathbf{z}_2)$ of corrupting one into the other in the transmission process through a communication channel is high.

Latent classes are placed on the nodes of a hierarchical star (Figure 1 (a)). Star can be completely described by a triple (L, \mathbf{b}, γ) , where L is the number of levels of the star, $\mathbf{b} = (b_1, \dots, b_{L-1})$ is a vector of branching degrees b_l for nodes on non-leaf levels $l = 1, \dots, L - 1$. $\gamma \in (0, 1)$ is a parameter determining the strength of the connection between two connected nodes (channel noise). The channel noise probabilities $P(\mathbf{z}_1|\mathbf{z}_2)$ are shown in Figure 1 (b). Probabilities between non-connected nodes are 0.

It is convenient to work with two copies \mathcal{Z}_Y and \mathcal{Z}_Z of the latent space \mathcal{Z} . For each user $\mathbf{u} \in \mathcal{U}$, the film-conditional ratings \mathbf{v} (type I) or pairs (\mathbf{v}, \mathbf{y}) (type II) are generated as follows:

- 1 randomly generate a latent class index $\mathbf{z}_Y \in \mathcal{Z}_Y$ by sampling the user-conditional probability distribution $P(\cdot|\mathbf{u})$ on \mathcal{Z}_Y .
- 2 transmit the class identification \mathbf{z}_Y through a noisy communication channel, and receive (a possibly different) class index $\mathbf{z}_Z \in \mathcal{Z}_Z$ with probability $P(\mathbf{z}_Z|\mathbf{z}_Y)$.
- 3 randomly generate a film-conditional rating \mathbf{v} with probability $P(\mathbf{v}|\mathbf{y}, \mathbf{z}_Z)$ (type I) or a pair (\mathbf{v}, \mathbf{y}) with probability $P(\mathbf{v}, \mathbf{y}|\mathbf{z}_Z)$ (type II).

The models for user ratings have now the following form:

$$P(\mathbf{z}_Z|\mathbf{u}) = \sum_{\mathbf{z}_Y \in \mathcal{Z}_Y} P(\mathbf{z}_Z|\mathbf{z}_Y)P(\mathbf{z}_Y|\mathbf{u}), \tag{1}$$

$$P(\mathbf{v}|\mathbf{y}, \mathbf{u}) = \sum_{\mathbf{z}_Z \in \mathcal{Z}_Z} P(\mathbf{v}|\mathbf{y}, \mathbf{z}_Z)P(\mathbf{z}_Z|\mathbf{u}) \quad \text{[type I]} \tag{2}$$

$$P(\mathbf{v}, \mathbf{y}|\mathbf{u}) = \sum_{\mathbf{z}_Z \in \mathcal{Z}_Z} P(\mathbf{v}, \mathbf{y}|\mathbf{z}_Z)P(\mathbf{z}_Z|\mathbf{u}) \quad \text{[type II]} \tag{3}$$

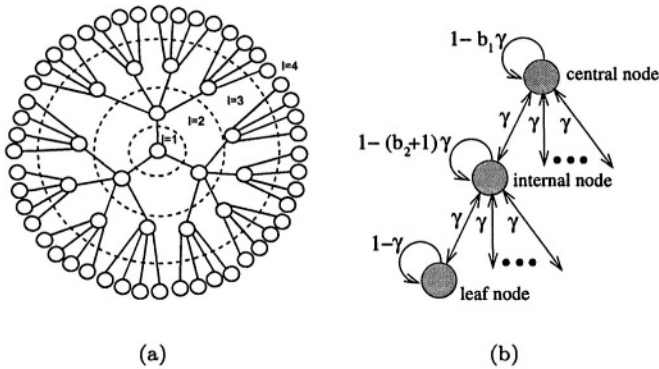


Figure 1. (a) An example of star topology of latent classes with four levels ($L = 4$). $b = (3, 4, 4)$. (b) Channel noise probabilities defining the star topology.

3. Parameter Estimation

Following [3], we denote by $\rho(v, y, z_Z)$ the probabilities $P(v|y, u)$ and $P(v, y|u)$ in type I and type II models, respectively.

To fit model parameters $P(z_Y|u)$ and $\rho(v, y, z_Z)$ to the observed data $\mathcal{D} = \{(u_1, y_1, v_1), \dots, (u_N, y_N, v_N)\}$, we use EM algorithm [1] that maximizes likelihood of data \mathcal{D} and iterates two steps - Expectation (E) and Maximization (M) - until convergence.

3.1 E-step

In the E-step, the algorithm computes the expected values of latent variables using the current values of the model parameters:

$$P(z_Y | y, u, v) = \frac{P(z_Y|u) \sum_{z_Z} \rho(v, y, z_Z) P(z_Z|z_Y)}{\sum_{z'_Y} P(z'_Y|u) \sum_{z_Z} \rho(v, y, z_Z) P(z_Z|z'_Y)}, \quad (4)$$

$$P(z_Z | y, u, v) = \frac{\rho(v, y, z_Z) \sum_{z_Y} P(z_Z|z_Y) P(z_Y|u)}{\sum_{z'_Z} \rho(v, y, z'_Z) \sum_{z_Y} P(z'_Z|z_Y) P(z_Y|u)}. \quad (5)$$

3.2 M-step

In the M-step, the algorithm re-estimates the model parameters by maximizing the expected complete data log-likelihood evaluated in the E-step. To derive the update equations, we need to determine the types of distributions for $P(z_Y|u)$ and $\rho(v, y, z_Z)$. It is natural to assume multi-

nomial $P(z_Y|u)$. However, for $\rho(v, y, z_Z)$ we use either multinomial, or binomial distribution. Multinomial distribution simply models probabilities of occurrence of ratings v . In the other hand, binomial distribution respects the ordering of rating values v^1 and imposes the assumption of unimodal rating distribution. We work with the following models:

- type I
 - I-Mult: multinomial $P(v|y, z_Z)$,
 - I-Bin: binomial $P(v|y, z_Z)$ (respects ordering of ratings).
- type II
 - II-Mult: joint multinomial $P(v, y|z_Z)$,
 - II-IndM: we assume conditional independence of v, y given z_Z : $P(v, y|z_Z) = P(v|z_Z)P(y|z_Z)$; both $P(y|z_Z)$ and $P(v|z_Z)$ are multinomials,
 - II-IndB: $P(v, y|z_Z) = P(v|z_Z)P(y|z_Z)$; $P(y|z_Z)$ is multinomially and $P(v|z_Z)$ is binomially distributed (respects ordering of ratings).

Update equation for $P(z_Y|u)$ is the same for all types of models:

$$P(z_Y|u) = \frac{\sum_{y \in \mathcal{Y}_u} P(z_Y | y, u, v_{u,y})}{|\mathcal{Y}_u|}. \tag{6}$$

3.2.1 I-Mult and II-Mult. Update equations are:

$$P(v | y, z_Z) = \frac{\sum_{u \in \mathcal{U}_{v,y}} P(z_Z | y, u, v)}{\sum_{v'} \sum_{u \in \mathcal{U}_{v',y}} P(z_Z | y, u, v')} \quad [\text{type I}] \tag{7}$$

$$P(v, y | z_Z) = \frac{\sum_{u \in \mathcal{U}_{v,y}} P(z_Z | y, u, v)}{\sum_{v',y'} \sum_{u \in \mathcal{U}_{v',y'}} P(z_Z | y', u, v')} \quad [\text{type II}] \tag{8}$$

where $\mathcal{U}_{v,y} = \{u \in \mathcal{U} | (u, y, v) \in \mathcal{D}\}$ is the set of users that evaluated film y with rating v .

3.2.2 II-IndM. Update equations are given by:

$$P(y|z_Z) = \frac{\sum_{u \in \mathcal{U}_y} P(z_Z | y, u, v_{u,y})}{\sum_{y'} \sum_{u \in \mathcal{U}_{y'}} P(z_Z | y', u, v_{u,y'})}, \tag{9}$$

¹i.e. it takes into account that ratings 4 and 5 are closer to each other than 1 and 5

$$P(v|z_Z) = \frac{\sum_y \sum_{u \in \mathcal{U}_{y,v}} P(z_Z | y, u, v)}{\sum_{v'} \sum_y \sum_{u \in \mathcal{U}_{y,v'}} P(z_Z | y, u, v')}. \quad (10)$$

3.2.3 II-IndB. Update equation for $P(y|z_Z)$ is the same as in eq. (9). $P(v|z_Z) = \binom{V}{v} p_{z_Z}^v (1 - p_{z_Z})^{V-v}$ is a binomial distribution with mean $p_{z_Z}|\mathcal{V}|$ and shape parameter p_{z_Z} . Update equation for parameter p_{z_Z} is given by:

$$p_{z_Z} = \frac{\sum_y \sum_{u \in \mathcal{U}_y} P(z_Z | y, u, v_{u,y}) v_{u,y}}{V \sum_y \sum_{u \in \mathcal{U}_y} P(z_Z | y, u, v_{u,y})}. \quad (11)$$

3.2.4 I-Bin. If $P(v|y, z_Z)$ is binomially distributed, then parameter $p_{z_Z,y}$ of the distribution is updated according to:

$$p_{z_Z,y} = \frac{\sum_{u \in \mathcal{U}_y} P(z_Z | y, u, v_{u,y}) v_{u,y}}{V \sum_{u \in \mathcal{U}_y} P(z_Z | y, u, v_{u,y})}. \quad (12)$$

4. Experiments

In this section we demonstrate latent class models with $K = 64$ latent classes and either star topology (STop) described by $(L = 4, \mathbf{b} = (3, 4, 4), \gamma = 0.15)$ or no topology (NTop) with 64 independent latent classes. Models of types I and II are trained with different distribution models for $\rho(v, y, z_Z)$, as described in Section 3.2.

We experimented with *EachMovie dataset*² containing ratings for films. User ratings are expressed on a 6-point scale. We selected a set of 100 most rated films. The number of users that rated at least one film from the selected set was 60, 895.

We partitioned the set of ratings into two sets – training and test sets. The training set \mathcal{D} is used to train the models and visualize the data. The test set T is used for evaluation of generalization capabilities of the models *within the set of users contained in \mathcal{D}* . Similarly to [3], we applied *all but one protocol*: 1 randomly selected rating from each user having at least 10 ratings was assigned to the test set.

The rating models are trained on the training set \mathcal{D} . To initialize model parameters we run SOM with the same star topology on the data \mathcal{D} (see [5]). After the initialization, models are trained with the EM algorithm³ and data are visualized. We use perplexity $P =$

²<http://www.research.compaq.com/SRC/eachmovie/>

³typically the data likelihood given by the models leveled up after 50 iterations

$\exp \left[-\frac{1}{|T|} \sum_{(u,y,v_{uv}) \in T} \log \rho(u,y,v) \right]$ of ratings on the training and test sets to compare the models.

4.1 Results

Tables 1 and 2 show perplexities of the models on training and test sets. For models of both types, perplexity of training data is smaller for multinomial distribution than for the binomial. In contrast, binomial distribution beats multinomial distribution on test data. This indicates that binomial distribution better regularizes the models by introducing less degrees of freedom (free parameters) and by imposing a unimodal structure on ordered rating values v . Models with unconstrained latent space (NTop) fit the training data better, but tend to overfit, as evidenced by better test set perplexities of models with imposed latent space topology (STop).

Table 1. Perplexity for training data.

	<i>I-Mult</i>	<i>I-Bin</i>	<i>II-Mult</i>	<i>II-IndM</i>	<i>II-IndB</i>
STop	2.86	3.62	159.28	157.79	191.04
NTop	2.64	3.50	134.90	140.18	170.70

Table 2. Perplexity for test data.

	<i>I-Mult</i>	<i>I-Bin</i>	<i>II-Mult</i>	<i>II-IndM</i>	<i>II-IndB</i>
STop	6.77	4.25	1636.92	1245.50	517.11
NTop	10.73	4.29	77243.87	33807.27	2353.47

4.2 Visualization

In this section we present some visualization plots of “common interest patterns” found by the models with star topology.

4.2.1 Models of Type II. Models of type II are suitable for visualizing the most probable films for each abstract class, i.e. films with largest $P(y|z) = \sum_v P(v,y|z)$. In order to understand to what degree are the films in particular classes similar, we present genre codes for latent classes in a hierarchical tree. Genre code for each class is made

up from genres of 5 most probable films for that class. Genres⁴ for films are taken from Internet Movie Database⁵ and they are represented by abbreviations shown in Table 3. The size of each genre abbreviation in genre codes for classes is proportional to the number of films (among the 5 most probable films) of that genre.

Table 3. Genre abbreviations.

A: Action	C: Comedy	D: Drama	E: Western	F: Family
H: Horror	L: Classic	M: Musical	N: Animation	P: Crime
R: Romance	S: Sci-Fi	T: Thriller	U: Mystery	V: Adventure
W: War	Y: Fantasy			

Figure 2 shows genre codes for classes of a model *STop II-IndB*. Topological organization of the classes according to the genres is clearly visible. The genre organization emerges naturally from the rating patterns in the data set and the imposed star (tree) topology on the latent classes.

4.2.2 Models of Type I. For models of type I, it is possible to visualize the rating distribution $P(v|\mathbf{y}, \mathbf{z})$ for each fixed film \mathbf{y} , given the latent class \mathbf{z} . By inspecting the latent-class-conditional rating distributions $P(\cdot|\mathbf{y}, \mathbf{z})$, we can demonstrate that “similar” films tend to have similar rating distributions. For illustration purposes, we choose four films: two are *romantic comedies* – *Ghost* and *Pretty Woman*, one film is a *criminal horror* – *Silence of the Lambs*, the last film is a *criminal drama* – *Pulp Fiction*.

Figure 3 visualizes rating patterns for those films as modeled by *STop I-Bin*. They are almost identical for the two *romantic comedies* (Figures 3(a) and 3(b)) and very different from *criminal drama* in Figure 3(c). The two *crime* films lead to both similar and dissimilar patterns on parts of the tree (Figures 3(c) and 3(d)).

For example, users concentrated in class 3 of the fourth level⁶ enjoy romantic comedies. In contrast, users represented by class 2 of the third level and classes 5-8, 25 and 27 of the fourth level tend to dislike them. Classes 2, 9, 20 and 33 of the fourth level represent users that like crimes. Class 22 on the fourth level represents a mixture of users that are attracted to both comedies and crimes.

⁴The film genres were not explicitly used in training the models.

⁵<http://www.imdb.com>

⁶All classes of each level are counted from the left to right.

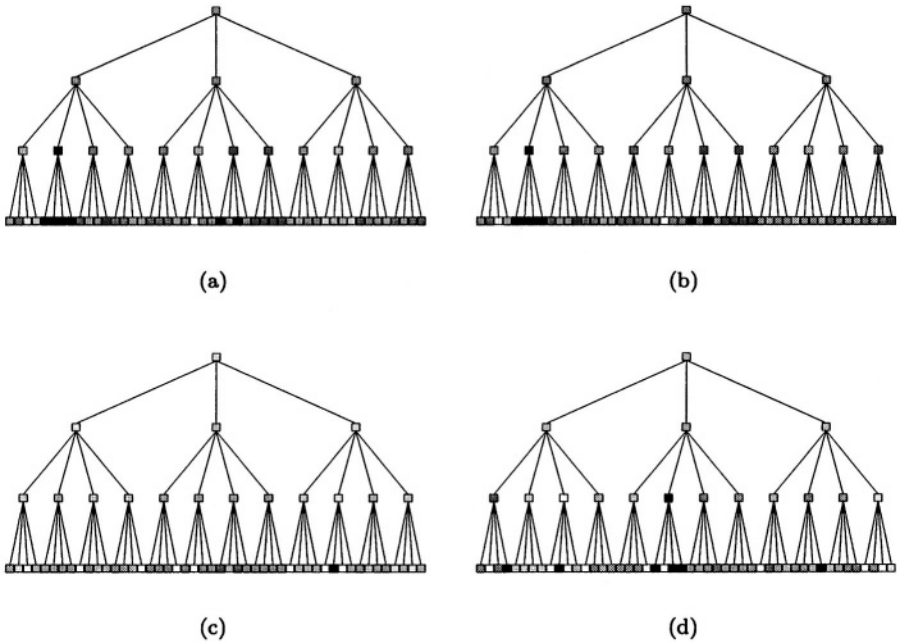


Figure 3. Model of type I – STop I-Bin. Shown are the means of the class-and-film-conditional rating distributions $P(v|y_*, z)$ for the following films: Ghost (a) Pretty Woman (b) The Silence of the Lambs (c) and Pulp Fiction (d). The brighter the color of the node is, the higher is its mean rating.

5. Conclusions

We have endowed two latent class models for collaborative filtering with star topology. Since our system is a probabilistic model of the data, we can use tools from data mining and probability and information theories to interpret the trends captured by the abstract latent classes. Topological organization of latent space makes orientation in rating patterns captured by latent classes easier and more systematic.

We demonstrated our system on a large collection of user ratings for films. We used different distribution to model (class conditional) variations in user ratings. For star topology, binomial distribution is more appropriate than multinomial, since it adds an extra degree of regularization by having free parameters and by imposing a unimodal structure on ordered set of rating values.

We plan to study construction of a hierarchy of topographic latent class models, where we would be able to interactively “zoom in” into interesting user groups and rating patterns.

Acknowledgments

This work was supported by the Slovak Grant Agency VEGA under under Grant No. VG 1/0162/03.

References

- [1] A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, B*, 39(1):1–38, 1977.
- [2] T. Hofmann. Probmap - a probabilistic approach for mapping large document collections. *Journal for Intelligent Data Analysis*, 4:149–164, 2000.
- [3] T. Hofmann. Learning what people (don't) want. In L. De Raedt and P. Flach, editors, *12th European Conference on Machine Learning (ECML)*, pages 214–225. Springer, 2001.
- [4] T. Hofmann, and J. Puzicha. Latent Class Models for Collaborative Filtering. *Proceedings of the International Joint Conference in Artificial Intelligence*, 1999.
- [5] **P. Tiño**, and **g. Polčicová**. Topographic organization of user preference patterns in collaborative filtering. *Neural Network World*, 13(3):311–324, 2003.