ITEM-USER PREFERENCE MAPPING WITH MIXTURE MODELS
Data Visualization for Item Preference

Yu Fujimoto
Dept. Integrated Information Technology, Aoyama Gakuin Univ., Kanagawa, Japan

Hideitsu Hino, Noboru Murata
Dept. Electrical Engineering and Bioscience, Waseda Univ., Tokyo, Japan

Keywords: Bradley-Terry model, Mixture model, EM algorithm, Preference map, Data visualization.

Abstract: In this paper, we propose a visualization technique of a statistical relation of users and preference of items based on a mixture model. In our visualization, items are given as points in a few dimensional preference space, and user specific preferences are given as lines in the same space. The relationship between items and user preferences are intuitively interpreted via projections from points onto lines. As a primitive implementation, we introduce a mixture of the Bradley-Terry models, and visualize the relation between items and user preferences with benchmark data sets.

1 INTRODUCTION

In a market research study or an item recommendation system, it is very important to model and interpret a statistical relation of "users" and "preference of items" based on a data set. Visualization of such models helps us to discover new relations between items and users, e.g. unknown preference tendency for a specific user. And visualization also supports to interpret results of item recommendation in systems.

It has an old history to model preference levels of items from the statistical aspect. Most researchers assume that a preference parameter θ_i is attached to the item T_i for I different items (Bradley and Terry, 1952; Luce, 1959; Plackett, 1975). With this parameter, favorable or unfavorable items for users are intuitively interpreted. However, the preference parameter potentially has absurdity, because it is obtained based on data which reflect various users' average, and the one-dimensional preference assumption may cause a wrong interpretation. Then, an idea of multiple preferences, which is an assumption that a user evaluates an item comprehensively with some indices such as one's interest and credibility of the item, is naturally introduced. For representation of multiple preferences, some applications of mixture models are proposed (Croon and Luijkx, 1993; Murphy and Martin, 2003).

In this paper, we propose a visualization of relation between items and users to assist analysis of multiple preferences based on mixture models. In our visualization, items are mapped in a K-dimensional space associated with K preference coordinates and their levels on a user specific preference are shown as projections onto a line on the map. With this mapping, preference relation between items and user preferences can be visually interpreted.

This paper is composed as follows. In Section 2, a simple probability model with preference parameters, broadly known as the Bradley-Terry (BT) model, and its mixtures are introduced. In Section 3, an idea for preference mapping which visualizes the item preference is explained. We also mention differences among other visualization tools and our method. Section 4 shows experimental results of the item preference mapping based on mixtures of BT models. And Section 5 is devoted to concluding remarks.

2 PREFERENCE MODEL AND ITS MIXTURE

Let T_i be the i-th item where $i = 1, \ldots, I$. Here, a preference parameter set $\theta = \{\theta_1, \ldots, \theta_I\}$ is introduced to represent relative preference levels for I items. Statis-

Fujimoto Y., Hino H. and Murata N. (2009).
ITEM-USER PREFERENCE MAPPING WITH MIXTURE MODELS - Data Visualization for Item Preference.
In *Proceedings of the International Conference on Knowledge Discovery and Information Retrieval*, pages 105-111
DOI: 10.5220/0002274001050111

tical models denoted with such a preference parameter set are called *preference models* in this paper. We introduce a simple preference model, the BT model (Bradley and Terry, 1952), and its mixture.

2.1 Bradley-Terry Model

Assume that a user evaluates two items T_i and T_j with ratings $r_i, r_j \in \mathbf{N}$, and each user chooses a preferred item from T_i and T_j by comparing r_i and r_j. Let $T_i \succ T_j$ be the event $r_i > r_j$, which indicates that T_i is chosen in the comparison of T_i and T_j, and X_{ij} be a variable for the comparison result which takes one of $\{T_i \succ T_j, T_j \succ T_i\}$. In the BT model, the probability that "the item T_i is preferred in the comparison of T_i and T_j", denoted as $p(T_i \succ T_j; \theta)$, is given by

$$p(T_i \succ T_j; \theta) = \frac{\theta_i}{\theta_i + \theta_j} \quad (i \neq j), \tag{1}$$

where

$$\sum_{i=1}^{I} \theta_i = 1, \qquad \theta_i > 0 \quad (i = 1, \ldots, I).$$

Intuitively speaking, the item T_i which has larger θ_i is chosen more frequently, and θ indicates a set of preference levels which is common to N users.

Let $X^n = \{X^n_{ij} | 1 \leq i < j \leq I\}$ be all the paired comparisons in I items, compared by the n-th user. Under the assumption that each comparison is independent, the probability $\Pr(X^n = x^n) = p(x^n; \theta)$, where x^n is an observation from the n-th user, is given by

$$p(x^n; \theta) = \prod_{i \neq j} \left(\frac{\theta^i}{\theta^i + \theta^j} \right)^{c^n_{ij}}, \tag{2}$$

where c^n_{ij} is an indicator, that is

$$(c^n_{ij}, c^n_{ji}) = \begin{cases} (1,0) & (x^n_{ij} = T_i \succ T_j) \\ (0,1) & (x^n_{ij} = T_j \succ T_i) \\ (0,0) & (x^n_{ij} \text{ is missed}). \end{cases} \tag{3}$$

Note that $(c^n_{ij}, c^n_{ji}) = (0,0)$ indicates that the comparison x^n_{ij} is missed because the items T_i or/and T_j are not rated, or their ratings are the same[1].

With Eq.(2), the log likelihood of $x^{1:N} = \{x^1, \ldots, x^N\}$ is given as follows,

$$\begin{aligned} L(\theta) &= \sum_{n=1}^{N} \log p(x^n; \theta) = \sum_{n=1}^{N} \sum_{i \neq j} c^n_{ij} \log \frac{\theta_i}{\theta_i + \theta_j} \\ &= \sum_{i \neq j} c_{ij} \log \frac{\theta_i}{\theta_i + \theta_j}, \end{aligned} \tag{4}$$

[1] The modeling of paired comparison data with ties ($r_i = r_j$) also has a long history (Rao and Kupper, 1967; Davidson and Beaver, 1977; Joe, 1990; Kuk, 1995), though tied cases are neglected for simplicity in this paper.

where $c_{ij} = \sum_{n=1}^{N} c^n_{ij}$.

The maximum likelihood (ML) estimation procedure for the BT model, which achieves

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}} \sum_{i \neq j} c_{ij} \log \frac{\theta_i}{\theta_i + \theta_j}, \tag{5}$$

has been already discussed from several contexts, and some iterative estimation methods for the ML estimator $\hat{\theta}$ have been introduced (Hastie and Tibshirani, 1998; Huang et al., 2006). In this paper, the estimation is achieved by the following algorithm used in Huang et al.(2006).

Algorithm 1. Estimation of BT model.

---

input *pairwise comparison data* $x^{1:N}$.
initialize $t = 0$, *and choose an initial parameter* $\theta^{(0)}$.
repeat *until convergence*
 update

$$\theta_i^{(t+1)} \leftarrow \frac{\sum_{j \neq i} c_{ij}}{\sum_{j \neq i} \frac{c_{ij} + c_{ji}}{\theta_i^{(t)} + \theta_j^{(t)}}}, \tag{6}$$

  for $i = 1, \ldots, I$.
 normalize $\theta^{(t+1)}$ *and set* $t \leftarrow t + 1$.
output *converged parameter vector* θ.

2.2 Mixture Model

In the previous subsection, the BT model was explained. In this subsection, a mixture of BT models and its estimation method are introduced.

We assume that "users evaluate items based on K preference parameter sets with their own weights", and introduce a mixture model whose component respectively represents a preference from a different point of view. Under this assumption, the distribution of X^n is given by the mixture of preference models,

$$p(X^n) = \sum_{k=1}^{K} p(M_k) p(X^n | M_k), \tag{7}$$

where M_k is the k-th preference model with the parameter set $\theta^k = \{\theta^k_1, \ldots, \theta^k_I\}$ and $p(X^n|M_k)$ is given by Eq.(2) with the parameter set θ^k. And the log likelihood for $x^{1:N}$ is given as

$$L(\Theta) = \sum_{n=1}^{N} \log \sum_{k=1}^{K} p(M_k) p(x^n | M_k), \tag{8}$$

where $\Theta = \{\theta^1, \ldots, \theta^K\}$ is the set of parameters for the mixture.

Since the direct maximization of Eq.(8) is complex, we apply the EM algorithm (McLachlan and Krishnan, 1996) to estimate the mixture of BTs. The objective function, so-called the Q-function, for the EM estimation is defined as follows,

$$
\begin{aligned}
&Q(\Theta;\Theta^{(t)}) \\
&= \sum_{n=1}^{N}\sum_{k=1}^{K} p(M_k|x^n;\Theta^{(t)}) \log p(x^n|M_k;\Theta) \qquad (9)\\
&= \sum_{n=1}^{N}\sum_{k=1}^{K} p(M_k|x^n;\Theta^{(t)}) \sum_{i\neq j} c_{ij}^n \log \frac{\theta_i^k}{\theta_i^k+\theta_j^k} \\
&= \sum_{k=1}^{K}\sum_{i\neq j} w_{ij}^{k(t)} \log \frac{\theta_i^k}{\theta_i^k+\theta_j^k}, \qquad (10)
\end{aligned}
$$

where $w_{ij}^{k(t)} = \sum_{n=1}^{N} c_{ij}^n p(M_k|x^n;\theta^{k(t)})$. We call $p(M_k|x^n)$ user weight. With this function, a procedure for the EM estimation is denoted as follows.

Algorithm 2. The EM estimation.

---

input *pairwise comparison data* $x^{1:N}$.
initialize $t = 0$, *and choose an initial parameter* $\Theta^{(0)}$
repeat *until convergence*

E-step: *calculate user weight* $p(M_k|x^n;\Theta^{(t)})$ *of Eq.*(9) *by*

$$p(M_k|x^n;\Theta^{(t)}) = \frac{\prod_{i\neq j} \frac{c_{ij}^n \theta_i^{k(t)}}{\theta_i^{k(t)}+\theta_j^{k(t)}}}{\sum_{m=1}^{K}\prod_{i\neq j} \frac{c_{ij}^n \theta_i^{m(t)}}{\theta_i^{m(t)}+\theta_j^{m(t)}}}. \qquad (11)$$

M-step: *maximize the Q-function given by Eq.*(10) *with respect to* Θ, *that is equivalent to*

$$\theta^{k(t+1)} = \underset{\theta^k}{\operatorname{argmax}} \sum_{i\neq j} w_{ij}^{k(t)} \log \frac{\theta_i^k}{\theta_i^k+\theta_j^k} \qquad (12)$$

for all k, and set $t \leftarrow t+1$.
output *converged parameter vector* $\hat{\Theta}$.

Note that Eq.(12) is equivalent to Eq.(5) and the maximization is achieved based on Algorithm 1 by using $w_{ij}^{k(t)}$ instead of c_{ij}. As broadly known, the EM algorithm is the local maxima algorithm. In our experiments, therefore, models are estimated five times from randomly chosen initial parameters, and the one which achieves the highest Q-value is used.

For selection of the optimal K, the information criteria (Akaike, 1974; Barron et al., 1998) or cross validation (CV)(Hastie et al., 2001) are applicable. In Section 4, we use CV to select K.

3 PREFERENCE MAPPING

As defined in Section 2, preference models have parameters to represent preferences for I items. With

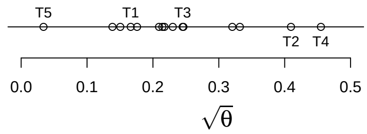

Figure 1: An example of preference mapping in one-dimensional case. Items are plotted according to the square root of θ to facilitate visualization.

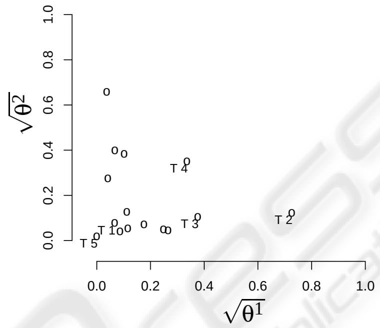

Figure 2: An example of 2D preference mapping.

a mixture of K preference models, an item T_i can be mapped on a K-dimensional space and its coordinate $(\theta_i^1,\ldots,\theta_i^K)$ gives us an intuitive interpretation of preference relation. In this section, we propose an intuitive preference mapping tool based on a mixture model.

3.1 Item Mapping

By introducing a preference model, the preference tendency of each item is interpreted intuitively by plotting the parameter on an axis. Figure 1 shows an example of such a plot. It is visually understood that T_4 tends to be preferred as shown in the figure. We call this type of plot a preference map[2].

In the case of a mixture model, θ^k shows one of the K coordinates for representation of preference levels. Accordingly, the multiple preference parameters with $K \geq 2$ can be mapped in a K-dimensional space in the same way. Figure 2 shows an example of two-dimensional preference mapping of $I = 15$ items. On a 2D preference map, those items which are commonly preferred by various users, like T_4, are mapped on the upper right side in the figure, and not preferred items in one dimension are mapped near the other axis, like T_2 and T_3.

[2]In this paper, all the axes in preference maps are given in the square root of preference parameters for easy-to-see visualization though it is not essential.

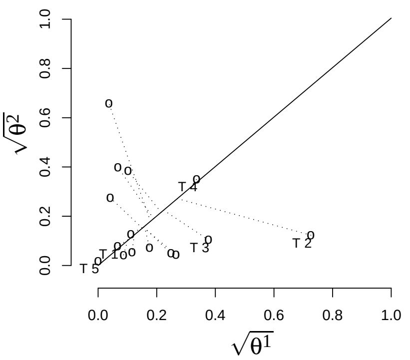

Figure 3: Preference mapping in a 2D space. A solid line in the map shows a user preference weight and dotted lines show projections onto the user preference levels which indicates ϕ^n.

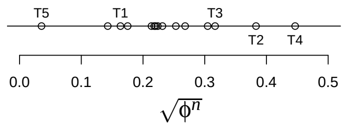

Figure 4: An example of a user preference ϕ^n. Plots are corresponding to the projected items onto the line in Figure 3.

3.2 User Preference

As previously denoted, we assume that "users evaluate items based on K preference parameter sets with their own weights" for the mixture model, and a user preference weight can be shown as a vector $(p(M_1|x^n),\dots,p(M_K|x^n))$ in a preference map. The direction of this vector expresses that the user thinks which coordinate is more important. For example, the solid line in Figure 3 shows a user preference weight for the user n which has tangents of $\frac{p(M_2|x^n)}{p(M_1|x^n)}$. The line indicates that the user rely on both of the axes θ^1 and θ^2.

Additionally, we can obtain a user specific preference levels $\{\phi_i^n\}$ of $\{T_i\}$ by projecting items onto the line on the map. Figure 3 also shows an example of item projections given as dotted lines to represent user preference levels. Projected points on the line indicate us the one-dimensional preference levels for the n-th user (see, Figure 4) embedded in the K-dimensional space. Such a projection can be expressed with the coordinate of T_i and the user preference weight. For example, one can easily come up with a projection given by the simple mixture of K parameters: a level of T_i on the n-th user preference is given by

$$\phi_i^n = \sum_{k=1}^{K} p(M_k|x^n)\theta_i^k. \tag{13}$$

Projection defined by Eq.(13) is called *linear projection* in this paper. For another example, a projection given by the mixture of K parameters in a sense of the single BT estimation is also possible, that is given by

$$\phi^n = \underset{\phi^n}{\operatorname{argmax}} \sum_{j\neq l} \left(\sum_{k=1}^{K} p(M_k|x^n) \frac{\theta_j^k}{\theta_j^k+\theta_l^k} \right) \log \frac{\phi_j^n}{\phi_j^n+\phi_l^n}, \tag{14}$$

where $\phi^n = \{\phi_1^n,\dots,\phi_I^n\}$. We call projection defined by Eq.(14) *BT projection*. The former type of projection is very simple however the latter one seems natural. In Figures 3 and 4, BT projection is applied to a specific user preference. In Section 4, two types of projections are compared.

As denoted at the end of previous section, the optimal K for the mixture model can be selected by the model selection procedure. However we still have a problem for visualizing preference mapping of a specific user. A visualization of mapped items is quite simple in the low dimensional case, however, in the case of $K \geq 4$, we have to visualize the high dimensional preference map in the low dimensional space. Here, user weight $p(M_k|x^n)$ shows that how much the n-th user emphasizes the k-th preference parameter set θ^k. An informative low dimensional preference map for the user is obtained by picking up some dimensions with the heaviest user weights. Note that the preference map without visualizing dimensions which have heavy user weights sometimes provides us misleading information. For example, Figure 5 shows an informative 2D preference map for a user by picking up θ^2 and θ^3 which have the two heaviest user weights in the mixture with $K = 4$, Figure 6 shows a misleading 2D preference map for the same user, mapped in a randomly selected two-dimensional space and Figure 7 shows the user preference levels ϕ^n. In Figure 5, the mapped items according to preference map coordinates θ^2 and θ^3 roughly reflect the user preference (ϕ^n), e.g. items mapped on the upper right side, like T_4, are also highly preferred in Figure 7. By contrast, items mapped on the upper right side in a misleading preference map do not preserve this relation, e.g. T_2 in Figure 6 is less preferred than T_4 in ϕ^n. Such a conflict happens because the low dimensional space with small $p(M_k|x^n)$ does not have valid information about the n-th user.

3.3 Other Mapping Tools

As a tool for visualization of probabilistic relation between evaluated items and users, the Multi-Dimensional Scaling (MDS) or the Multi-Dimensional unfolding (Bennett and Hays, 1960) has been proposed. Especially the Multi-Dimensional unfolding is a method to map items and users in a low

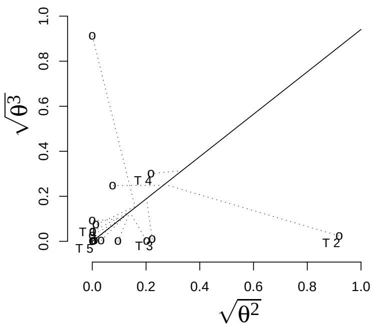

Figure 5: An example of informative 2D preference mapping of the mixture with $K = 4$.

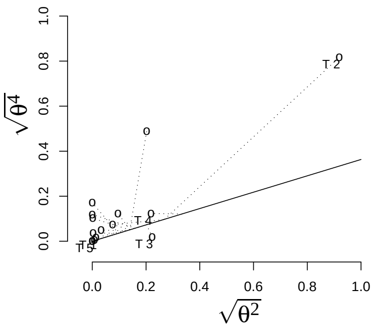

Figure 6: An example of misleading preference mapping.

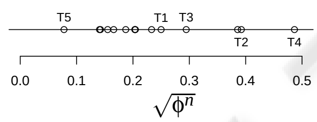

Figure 7: User preference levels ϕ^n corresponding to the lines in Figure 5 and Figure 6.

dimensional space according to user preference levels of items. Even in recent years, some researchers also have proposed the preference visualization (Mei and Shelton, 2006; Zenebe and Norcio, 2007). An obvious difference between their works and ours is the representation of user preference in the map. In our method, items are mapped in the K-dimensional space and their preference levels for each user are given by projections onto the corresponding line whose direction indicates the user weight: i.e. how much this user relies on the axes of the graph relatively. We avoid to map users and items as points on the same space because they are not the same kind of data.

To visualize relations between items and user preferences with points and lines, the Arrow and Point Method (APM) (Hayashi, 1993) has been also proposed. The APM achieves similar mapping as our method, and also shows the relation between items and users by points and lines. However the map obtained by our method shows K universal preference coordinates at the same time and intuitively interpreted from a statistical viewpoint, which is a discriminative point that the APM does not have.

4 EXPERIMENTS

In this section, we discuss our visualization method on real-world data sets. At first, we show the experimental results of tuning the optimal K in a way of the conventional statistical model selection. Then, we evaluate a structure of the preference mapping which is defined by a mixture of preference models from a viewpoint of precision of user preferences with a ranking correlation metric.

4.1 Real-world Data Sets

The MovieLens data set is a standard benchmark data set provided by GroupLens research team (Riedl and Konstan, 2000). This data set contains $100,000$ ratings answered by $N = 943$ users for $I = 1,682$ movies.

The BookCrossing data set is another benchmark data set which contains ratings of books (Ziegler et al., 2005). To obtain dense paired comparison set for evaluating our preference mapping, we removed answers rated as 0 and picked up $4,282$ ratings answered by $N = 799$ users for $I = 200$ items.

We use these two data sets for evaluation of our mixture model and preference mapping.

4.2 Evaluation Metric

For a metric of the mapping precision, we use the ranking correlation between rating data and user preference levels. Let r_i^n and r_j^n be ratings of T_i and T_j evaluated by the n-th user. For items evaluated by the user, if the ordering of r_i^n and r_j^n is the same as the ordering of ϕ_i^n and ϕ_j^n, a pair (T_i, T_j) is called concordant. On the other hand, if the two orderings are different, the pair is called discordant. In the case of ties, that is $r_i^n = r_j^n$ (or/and $\phi_i^n = \phi_j^n$), the pair is neither concordant or discordant. Then the metric is given as follows

$$\tau_b^n = \frac{N_C^n - N_D^n}{\sqrt{N_C^n + N_D^n + N_r^n}\sqrt{N_C^n + N_D^n + N_\phi^n}}, \qquad (15)$$

where N_C^n (N_D^n) is the number of concordant (discordant) pairs, and N_r^n (N_ϕ^n) is the number of ties in ratings (preference levels) of the n-th user. Equation (15) is called Kendall's tau-b, and used for evaluating the correlation between two ranking sequences with ties (Mei and Shelton, 2006). Note that the tau-b metric

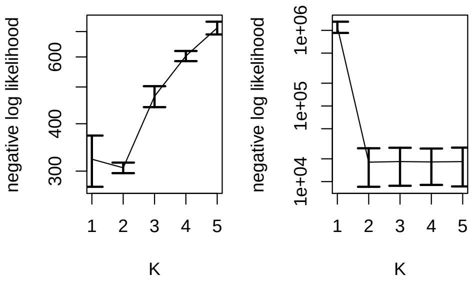

Figure 8: Plots of 5-fold CV results of the negative log likelihood: left one is the result of BookCrossing data and right one is that of MovieLens data. The vertical bars in the graphs show standard deviations.

is $-1 \leq \tau_b^n \leq 1$, and $\tau_b^n \cong 1$ ($\tau_b^n \cong -1$) indicates the orderings of the ratings and the user preferences are strongly positive (negative) correlated. And $\tau_b^n = 0$ indicates their orderings are independent. Note that the tau-b metric is calculated between all the pairs of ratings and user preference levels such that $c_{ij}^n = 1$ for each user. In the next subsection, we experimentally show that the mixture of preference models achieves precise mapping in the viewpoint of the tau-b metric.

4.3 Experimental Results

To visualize preference maps of the BookCrossing data and the MovieLens data, the optimal K is selected based on 5-fold CV. The results (Figure 8) show that $K = 2$ is selected for both of the BookCrossing data and the MovieLens data in the sense of the one standard deviation rule. Figure 9 shows simplified distributions of tau-b values between ratings and user preferences, defined by Eq.(15), of N users. A user preference ϕ^n is calculated under the optimal K selected by CV (i.e. $K = 2$) and two types of projections (linear and BT projections) are applied to derive ϕ^n. And the tau-b values between ratings and θ in the single BT model is also plotted in Figure 9 for comparison. In the figure, mixtures with the both projections apparently improve tau-b values and the results show that the mixture organizes more precise preference than the single preference model for each user. In other words, ranking of ratings corresponds to that of ϕ^n on each user preference more specifically by introducing mixture of preference models. The figure also shows that precision of ranking estimation with user preference based on the simple linear projection is approximately the same as that of the BT projection.

Figure 10 shows 2D preference maps as a result of the experiment on the BookCrossing data set. The left figure is the map with a user preference of those who rely on θ^1. For such a user, items which have a higher

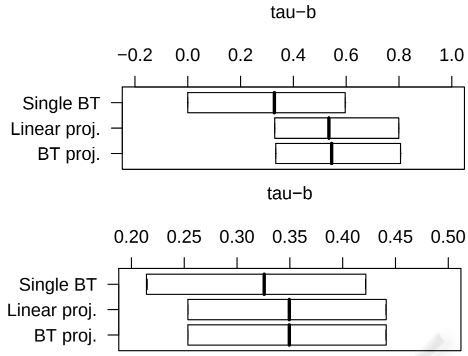

Figure 9: Plots of tau-b values. The line in each box indicates the median of tau-b for N users, and the vertical edges of the box indicate the 75% and 25% quantiles.

value in θ^1, like T_{171} shown as a filled square in the figure, are attractive (see the upper right figure). On the other hand, the middle figure is the map with the index of those who rely on θ^2. For such a user, an item like T_{55} shown as a filled circle is preferred (see the lower right). As the result, we can say that preference maps like Figure 10 has a potential of visualization which indicates not only relations between items but also differences between each user preference.

5 CONCLUSIONS AND FUTURE WORKS

This paper proposes a method to visualize multiple preferences of items, with a mixture of preference models. We adopt BT models for simple implementation of preference model in this paper, while the mixture can be composed of not only BT models but also any probability models for preference levels. Actually, as described in Hastie and Tibshirani (1998), the conventional BT model sometimes leads to the inaccurate preference parameter θ which does not reflect the preference order of items. To avoid this problem, a mixture of modified BT models, such as Huang et al. (2006), or alternative preference models (Luce, 1959; Plackett, 1975; Hino et al., 2009) should be applied. We also confirmed that a preference map can be drawn with other preference models though it doesn't explain in this paper. Regardless of whether BT models are used, the obtained map based on proposed method is directly interpreted as the probability model, and provides effective suggestions as the analysis result, e.g. we can show grounds of the recommendation visually for users.

In our method, items are mapped in the K-dimensional space, user's preference weights are

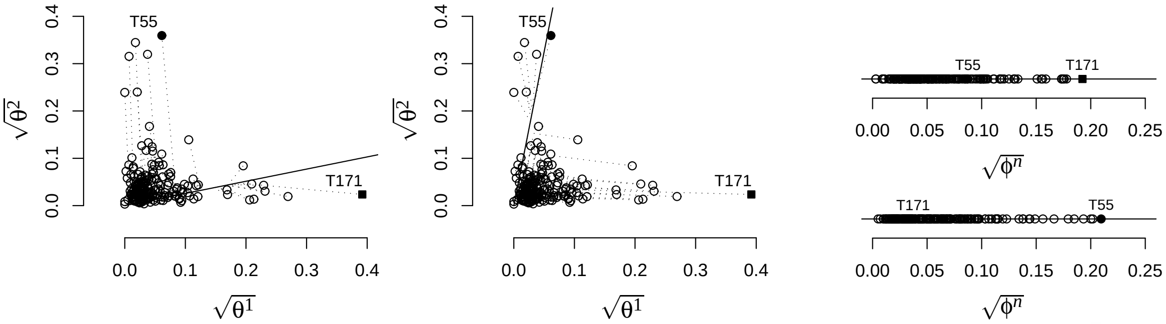

Figure 10: Preference maps based on the BookCrossing data set. The left figure shows a user preference of those who rely on θ^1, the middle one shows that of those who rely on θ^2 and the right figures show user preference levels ϕ^n of these users.

given as lines, and user's own preference levels are given by projections onto the corresponding line. We experimentally compared two types of projections, linear and BT, and verified that there was no big difference in results of tau-b metric. As a criterion to visualize the K-dimensional preference map and a user preference in a low dimensional space, we focus on the user weight $p(M_k|x^n)$. However, when we compare two users, the low dimensional map which accurately shows difference between their preferences is expected, though it remains as a future work.

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Automatic Control*, 19(6):716–723.

Barron, A., Rissanen, J., and Yu, B. (1998). The minimum description length principle in coding and modeling. *IEEE Trans. Information Theory*, 44(6):2743–2760.

Bennett, J. F. and Hays, W. L. (1960). Multidimensional unfolding: Determining the dimensionality of ranked preference data. *Psychometrika*, 25(1):27–43.

Bradley, R. A. and Terry, M. E. (1952). Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika*, 39(3/4):324–345.

Croon, M. and Luijkx, R. (1993). *Probability Models and Statistical Analyses for Ranking Data*, chapter 4, pages 53–74. Lecture Notes in Statistics. Springer.

Davidson, R. R. and Beaver, R. J. (1977). On extending the Bradley-Terry model to incorporate within-pair order effects. *Biometrics*, 33(4):693–702.

Hastie, T. and Tibshirani, R. (1998). Classification by pairwise coupling. *Annals of Statistics*, 26(2):451–471.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. 2nd Ed. Springer Series in Statistics. Springer.

Hayashi, C. (1993). Quantitative social research -belief systems, the way of thinking and sentiments of five nations -. *Behaviormetrika*, 19(2):127–170.

Hino, H., Fujimoto, Y., and Murata, N. (2009). Item preference parameters from grouped ranking observations. In *Proc. 13th Pacific-Asia Conference on KDD*.

Huang, T.-K., Weng, R. C., and Lin, C.-J. (2006). Generalized Bradley-Terry models and multi-class probability estimates. *Journal of Machine Learning Research*, 7:85–115.

Joe, H. (1990). Extended use of paired comparison models, with application to chess rankings. *Applied Statistics*, 39(1):85–93.

Kuk, A. Y. C. (1995). Modeling paired comparison data with large numbers of draws and large variability of draw percentages among players. *The Statistician*, 44(4):523–528.

Luce, R. D. (1959). *Individual Choice Behavior*. John Wiley & Sons, Inc.

McLachlan, G. J. and Krishnan, T. (1996). *The EM Algorithm and Extensions*. Wiley-Interscience.

Mei, G. and Shelton, C. R. (2006). Visualization of collaborative data. In *Proc. 22nd International Conference on UAI*, pages 341–348. AUAI Press.

Murphy, T. B. and Martin, D. (2003). Mixtures of distance-based models for ranking data. *Computational Statistics & Data Analysis*, 41:645–655.

Plackett, R. L. (1975). The analysis of permutations. *Applied Statistics*, 24(2):193–202.

Rao, P. V. and Kupper, L. L. (1967). Ties in paired-comparison experiments: A generalization of the Bradley-Terry model. *Journal of the American Statistical Association*, 62(317):195–204.

Riedl, J. and Konstan, J. (2000). Movielens data set. http://www.grouplens.org/.

Zenebe, A. and Norcio, A. F. (2007). Visualization of item features, customer preference and associated uncertainty using fuzzy sets. In *Proc. Annual Meeting of the North American Fuzzy Information Processing Society*, pages 22–32.

Ziegler, C.-N., McNee, S. M., and Konstan, J. A. (2005). Improving recommendation lists through topic diversification. In *Proc. 14th International World Wide Web Conference (WWW '05)*, pages 22–32.