Efficient Learning of User Conformity on Review Score

Kazumi Saito[1], Kouzou Ohara[2], Masahiro Kimura[3], and Hiroshi Motoda[4,5]

[1] School of Administration and Informatics, University of Shizuoka
k-saito@u-shizuoka-ken.ac.jp
[2] Department of Integrated Information Technology, Aoyama Gakuin University
ohara@it.aoyama.ac.jp
[3] Department of Electronics and Informatics, Ryukoku University
kimura@rins.ryukoku.ac.jp
[4] Institute of Scientific and Industrial Research, Osaka University
motoda@ar.sanken.osaka-u.ac.jp
[5] School of Computing and Information Systems, University of Tasmania

Abstract. We propose a simple and efficient method that learns and assesses the conformity of each user of an online review system from the observed review score record. The model we use is a modified Voter model that takes account of the conformity of each user. Conformity is learnable quite efficiently with a few tens of iterations by maximizing the log-likelihood given the observed data. The proposed method was evaluated and confirmed effective by two review datasets. It could identify both high and low conformity users. Users with high conformity are not necessarily early adopters. Their scores are influential to drive the consensus score. The user ranking of conformity was compared with PageRank and HITS in which user network was roughly approximated by the directed graph induced by the observed data. The proposed method gives more interpretable ranking, and the global property of high conformity users was identified.

Keywords: Social media, Conformity, Review score, Learning

1 Introduction

The emergence of Social Media has provided us with the opportunity to collect a large number of user reviews for various items, e.g., products and movies. People now read these reviews and make decisions about their actions, e.g. buy the product, see the movie. It is thus important to be able to assess the review influence of each user who posted a review. As such influence, we focus on the conformity defined as a type of social influence involving a change in belief or behavior in order to fit in with a group. Analyzing review in depth needs natural language processing. Fortunately many social media sites offer review scores, i.e. rating score which is a numeric value, for individual items as well as the user reviews. We use this review score instead of review itself.

The scores are rated by many users and their values vary across users and items. We want to identify high-quality items in a given category in an efficient way from these review scores. Naive ranking would be simply to rank items according to the number of reviews or the average review score. There the emphasis is more on the statistical reliability and not on the quality and reputation of each review. They do not account for the review influence of each user that rated a specific item. Several researchers [7, 6, 3, 8] incorporated the information of trust relationships with trust strengths (i.e., the local trust-value information) into low-rank matrix factorization techniques [10, 5], and improved the performance in rating prediction. Tang et al. [12] proposed a method of incorporating the information of trustworthiness of users (i.e., the global trust-value information) into this framework, and further improved the performance in rating prediction. They measured the trustworthiness of a user by applying the PageRank algorithm [1] to the trust network.

However, these existing techniques assume a separate source of information, often represented by a trust network, which is not available except for a relatively limited number of review sites. In addition, it is quite difficult to precisely obtain the trust network due to its intrinsic time varying nature. In contrast, we can easily observe the phenomena related to conformity (or herding) of people, as studied in [9], and it is possible to estimate the conformity metric of each user based on time series data of review scores for items given by users, that is, the data consisting of 4-tuple (u, i, s, t) which means user u rated item i and gave score s. The conformity metric proposed in this paper is so designed that users with high conformity tend to lead users' rating behavior for those items they rate, and users with low conformity tend to give unusual ratings for them.

We propose a simple and efficient method that learns and assesses the conformity of each user from the observed record of review scores without a need of separate trust network. The model we use is a modified version of Voter model in which we introduce the conformity of each user as a new variable to learn under the assumption that each user rates an item only once and the score is approximated by a multinomial distribution. The Voter model is one of the simplest models of opinion formation and propagation [11, 2]. It assumes that a user updates her opinion based on her own and her direct neighbors' opinions, i.e. following the majority similarly to conformity. We assume that a user rates an item only once and never update the value, which is different from the basic assumption of Voter model, but we borrows the idea that a user read other users' reviews and makes her choice probabilistically.

The conformity is learned such that the users' review score distributions predicted by the generative modified Voter model best match the observed score distributions, i.e. by maximizing the log-likelihood. The learning uses EM-like iterative scheme and it is very efficient taking full advantage of the convexity of the auxiliary objective function. Data are divided into half and the former half is used for learning the parameters of the model, i.e. conformity metrics, and generalization capability of the learned model is evaluated by the likelihood estimated by the unseen latter half data to determine the optimal value of the regularization factor that is introduced to avoid overfitting. Once the regularization factor is fixed, all the data are used to relearn the model.

We tested the proposed method by applying it to two review systems, Cosmetics review dataset[1] and Anime review dataset[2]. The conformity metric distribution of users depends on the characteristics of dataset used. What the proposed method found is that in the cases of adequate regularization factors, the majority of the people have the aver-

[1] http://www.cosme.net/
[2] http://www.anikore.jp/

age conformity metric and only a small fraction of people have high or low conformity metrics. Thus, the method can identify two interesting groups of people that is worth paying attention to, one with high conformity metrics and the other with low conformity metrics. Users with high conformity metrics are not necessarily early adopters. Their scores are influential to drive the expected consensus score, i.e. lead global behaviors. Users with low conformity metrics deviates from the average behavior. High conformity user has the following properties: 1) she rates many items, 2) there are many followers of her who rate the same items that she rated and 3) the scores of the followers are similar to her scores. These properties are quite natural and the proposed method confirmed them. From an analogy that the trustworthiness of a user is estimated by applying the PageRank algorithm to the trust network, we have compared the ranking results of PageRank and HITS algorithms if they have the same properties. To run these algorithms, we constructed the user network by creating a directed network for each item based on the time stamp information of the observed data. The results of Page Rank and HITS are not as clear as the proposed method.

2 Model

We denote the sets of users and items by $V = \{u, v, w, \cdots\}$ and $I = \{i, \cdots\}$, respectively. When a user $v \in V$ reviewed an item $i \in I$, we denote its timestamp and score by $t_{v,i}$ and $s_{v,i}$, respectively, where each score $s_{v,i}$ is denoted by a positive integer in $S = \{1, \cdots, |S|\}$, and $|S|$ stands for the number of elements in S. Then, we can express our observed data set as $D = \{\cdots, (v, i, t_{v,i}, s_{v,i}), \cdots\}$. Hereafter, let $V(i) = \{v \mid (v, i, t_{v,i}, s_{v,i}) \in D\}$ be a set of users who reviewed an item i. For users in $V(i)$, let $U(i,t) = \{u \in V(i) \mid t_{u,i} < t\}$ be the set of users whose review times are before t, and $U(i,t,s) = \{u \in U(i,t) \mid s_{u,i} = s\}$ the set of those users whose review score is s.

As mentioned earlier, users may decide their review scores of each item by taking account not only of their own evaluations, but also of past majority scores or those submitted by high conformity users. In order to stochastically cope with the opinion decision problem affected by majority scores, we can employ the basic voter model, and define the probability that a user v gives a score s to an item i at time t as $P_0(s \mid i,t) = (1+|U(i,t,s)|)(|S|+|U(i,t)|)$, where we employed a Bayesian prior known as the Laplace smoothing. Here we note that the Laplace smoothing corresponds to the assumption that each node initially holds one of the $|S|$ scores with equal probability. Note also that the Laplace smoothing corresponds to a special case of Dirichlet distributions that are very often used as prior distributions in Bayesian statistics.

Thus far, we assumed that all the past user scores are equally weighted. However, it is naturally conceivable that some high conformity users should have larger weights. In order to reflect this kind of effects into the model, we consider introducing a positive conformity metric $\exp(\theta_u)$ to each user u, where θ_u is a parameter which controls the conformity metric. Hereafter, we denote the vector consisting of these parameters by $\boldsymbol{\theta} = (\cdots, \theta_u, \cdots)$. Then, we can extend the basic Voter model $P_0(s \mid i,t)$ and build a generative model in which user v gives score s for item i at time t with the following probability.

$$P(s \mid i, t;\ \boldsymbol{\theta}) = \frac{1 + \sum_{u \in U(i,t,s)} \exp(\theta_u)}{|S| + \sum_{u \in U(i,t)} \exp(\theta_u)}. \tag{1}$$

In this paper, in order to estimate $\boldsymbol{\theta}$ from the observed data set D, we consider maximizing the following logarithmic likelihood function based on Eq. (1):

$$L(D;\ \boldsymbol{\theta}) = \sum_{i \in I} \sum_{v \in V(i)} \log P(s_{v,i} \mid i, t_{v,i};\ \boldsymbol{\theta}). \tag{2}$$

Here each review score s is replaced by the observed pair of $t_{v,i}$ and $s_{v,i}$ in Eq. (2).

3 Learning Algorithm

The number of users, which corresponds to the dimensionality of $\boldsymbol{\theta}$, easily becomes quite large, say tens of thousands. Thus, in order to avoid an overfitting problem, we consider minimizing the objective function with a standard regularization term defined by $J(\boldsymbol{\theta}) = -L(D\ ;\ \boldsymbol{\theta}) + \frac{\eta}{2}\|\boldsymbol{\theta}\|^2$, where $\|\boldsymbol{\theta}\|^2 = \sum_{u \in V} \theta_u^2$, and η stands for a regularization factor. In order to minimize $J(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$, we propose a learning algorithm based on the gradient descend method equipped with the second-order optimal step-length calculation for an auxiliary objective function of $J(\boldsymbol{\theta})$.

Now, let $I(v) = \{i \mid (v, i, t_{v,i}, s_{v,i}) \in D\}$ be a set of items which were reviewed by a user v. For users $V(i)$, let $W(i,t) = \{w \in V(i) \mid t_{w,i} > t\}$ be the set of users whose review time is after t and $W(i,t,s) = \{w \in W(i,t) \mid s_{w,i} = s\}$ be the set of those users whose review score is s. Moreover, we define two terms, $q_{v,i,t}(\boldsymbol{\theta}) = \exp(\theta_v)/(|S| + \sum_{u \in U(i,t)} \exp(\theta_u))$ and $q_{v,i,t,s}(\boldsymbol{\theta}) = \exp(\theta_v)/(1 + \sum_{u \in U(i,t,s)} \exp(\theta_u))$, just like posterior probabilities typically used in the EM algorithm. Then, by setting the search direction $\boldsymbol{\delta} = (\cdots, \delta_v, \cdots)$ to the negative gradient, i.e., $\delta_v = -\partial J(\boldsymbol{\theta})/\partial \theta_v$, we can calculate δ_v as follows:

$$\delta_v = \sum_{i \in I(v)} \sum_{w \in W(i, t_{v,i}, s_{v,i})} q_{v,i,t_{w,i},s_{v,i}}(\boldsymbol{\theta}) - \sum_{i \in I(v)} \sum_{w \in W(i, t_{v,i})} q_{v,i,t_{w,i}}(\boldsymbol{\theta}) + \eta \theta_v. \tag{3}$$

From the first term of the right-hand-side in Eq. (3), we can easily see that θ_v of user v tends to increase if $|W(i, t_{v,i}, s_{v,i})|$ is relatively large for many items i.

By using two terms defined by $N(v, i;\ \boldsymbol{\theta}) = \log(1 + \sum_{u \in U(i, t_{v,i}, s_{v,i})} \exp(\theta_u))$ and $D(v, i;\ \boldsymbol{\theta}) = \log(|S| + \sum_{u \in U(i, t_{v,i})} \exp(\theta_u))$, we can rewrite our objective $J(\boldsymbol{\theta})$ as follows:

$$J(\boldsymbol{\theta}) = -\sum_{i \in I} \sum_{v \in V(i)} N(v, i;\ \boldsymbol{\theta}) + \sum_{i \in I} \sum_{v \in V(i)} D(v, i;\ \boldsymbol{\theta}) + \frac{\eta}{2} \sum_{v \in V} \theta_v^2.$$

By using a term defined by $M(v, i;\ \boldsymbol{\theta} \mid \bar{\boldsymbol{\theta}}) = \sum_{u \in U(i, t_{v,i}, s_{v,i})} \theta_u\ q_{u,i,t_{v,i},s_{v,i}}(\bar{\boldsymbol{\theta}})$, we consider the following auxiliary objective function $Q(\boldsymbol{\theta} \mid \bar{\boldsymbol{\theta}})$ to describe our method for stably calculating the step-length λ with respect to $\boldsymbol{\delta}$.

$$Q(\boldsymbol{\theta} \mid \bar{\boldsymbol{\theta}}) = -\sum_{i \in I} \sum_{v \in V(i)} M(v, i;\ \boldsymbol{\theta} \mid \bar{\boldsymbol{\theta}}) + \sum_{i \in I} \sum_{v \in V(i)} D(v, i;\ \boldsymbol{\theta}) + \frac{\eta}{2} \sum_{v \in V} \theta_v^2.$$

where $\bar{\boldsymbol{\theta}}$ stands for the current estimate of $\boldsymbol{\theta}$. Here, by introducing a virtual node $z \notin V$ whose parameter value is fixed at zero, i.e., $\theta_z = 0$, an augmented set defined by $UA(i, s, t) = U(i, s, t) \cup \{z\}$ for each pair of item i and score s, and a cross-entropy term, $H(v, i;\ \boldsymbol{\theta} \mid \bar{\boldsymbol{\theta}}) = -\sum_{u \in UA(i, t_{v,i}, s_{v,i})} q_{v,i,t_{u,i},s_{u,i}}(\bar{\boldsymbol{\theta}}) \log q_{v,i,t_{u,i},s_{u,i}}(\boldsymbol{\theta})$, we can obtain $N(v, i;\ \boldsymbol{\theta}) = M(v, i;\ \boldsymbol{\theta} \mid \bar{\boldsymbol{\theta}}) + H(v, i;\ \boldsymbol{\theta} \mid \bar{\boldsymbol{\theta}})$. Thus, we can see that the objective function $J(\boldsymbol{\theta})$ is expressed as $J(\boldsymbol{\theta}) = Q(\boldsymbol{\theta} \mid \bar{\boldsymbol{\theta}}) - \sum_{i \in I} \sum_{v \in V(i)} H(v, i;\ \boldsymbol{\theta} \mid \bar{\boldsymbol{\theta}})$. Therefore, since $H(v, i;\ \boldsymbol{\theta} \mid \bar{\boldsymbol{\theta}})$ is minimized at $\boldsymbol{\theta} = \bar{\boldsymbol{\theta}}$, the objective function $J(\boldsymbol{\theta})$ is optimized by minimizing the auxiliary objective function $Q(\boldsymbol{\theta} \mid \bar{\boldsymbol{\theta}})$ with respect to $\boldsymbol{\theta}$.

By considering a univariate objective function defined as $F(\lambda) = Q(\bar{\boldsymbol{\theta}} + \lambda \boldsymbol{\delta} \mid \bar{\boldsymbol{\theta}})$, we can calculate the second-order optimal step-length λ as $\lambda = -F'(0)/F''(0)$, where $F'(0)$ and $F''(0)$ means the first- and second-order derivatives of $F(\lambda)$ with respect to λ at $\lambda = 0$. Note that $F'(0) = -\|\boldsymbol{\delta}\|^2$, and we can efficiently calculate $F''(0)$ as follows:

$$F''(0) = \sum_{i \in I} \sum_{v \in V(i)} \sum_{u \in U(i, t_{v,i})} \delta_u^2\, q_{u,i,t_{v,i}}(\bar{\boldsymbol{\theta}}) - \sum_{i \in I} \sum_{v \in V(i)} \Big(\sum_{u \in U(i, t_{v,i})} \delta_u\, q_{u,i,t_{v,i}}(\bar{\boldsymbol{\theta}}) \Big)^2 + \eta \|\boldsymbol{\delta}\|^2.$$

We can easily see that $F''(0) > 0$, $\lambda > 0$ and $F(\lambda)$ has a unique global optimal solution due to $F''(0) > 0$. Namely, we can stably calculate the step-length without using a technique like safeguarded Newton's method. Note that straightforward univariate objective function defined by $G(\lambda) = J(\bar{\boldsymbol{\theta}} + \lambda \boldsymbol{\delta})$ does not hold this kind of nice properties.

4 Experiments

We collected review score records from two famous review sites in Japan and constructed two datasets for this experiment. One consists of review scores for cosmetics extracted from "@cosme"[3] which is a Japanese word-of-mouth communication site for cosmetics. We refer to this dataset as the Cosmetics review dataset. The other one is composed of review records collected from "anikore"[4], a ranking and review site for anime, which is referred to as the Anime review dataset. In both the datasets, each record has 4-triple (u, i, s, t) as mentioned above, which means user u gives a score s to item i at time t. The Cosmetics review dataset has $297,453$ review records by $10,403$ users for $46,398$ items from 2008/12/07 to 2009/12/09, while the Anime review dataset has $300,327$ records by $13,112$ users for $1,790$ items from 2010/8/01 to 2012/8/08. The score is an integer value ranging from 1 to 7 in the Cosmetics dataset and from 1 to 5 in the Anime dataset.

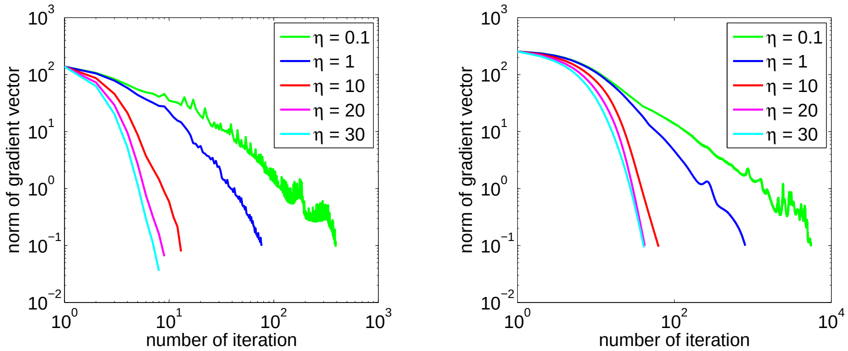
4.1 Learning Results

First of all, we evaluated how the regularization factor η affects the learning performance of the proposed method. To this end, we divided each dataset in half and learned the conformity metric of each user from the former half (training data) varying the value of η, and then evaluated the generalization capability of the learned model by using the latter half (test data) in terms of perplexity defined as $exp(-L(D_{test}; \hat{\boldsymbol{\theta}}))$ where D_{test}

[3] http://www.cosme.net/
[4] http://www.anikore.jp/

Table 1. Fluctuation of perplexity for the test data as a function of η.

η	0.1	1.0	10	20	30
Cosme	6.2073	6.0730	6.0404	6.0403	6.0408
Anime	2.8118	2.7685	2.7669	2.7678	2.7685



(a) Cosme dataset (b) Anime dataset

Fig. 1. Number of iterations spent by the proposed algorithm for different values of η.

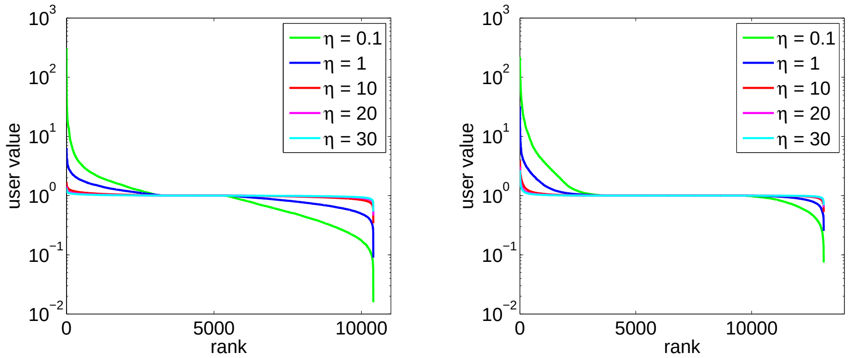
means the test data and $\hat{\boldsymbol{\theta}}$ a parameter vector learned from the training data. Table 1 shows the resulting values of the perplexity for different regularization factor η. Here, note that the smaller the value of perplexity is, the better a learned model is in terms of the generalization capability. From this result, we see that the resulting perplexity becomes smaller as η becomes larger, and levels off for η greater than a certain threshold, say 10 in these datasets. This means that the proposed method is robust for a wide rage of η and can achieve good generalization capability unless η is too small.

Next, we investigated how the value of η affects the efficiency of the proposed method by comparing the number of iterations spent by the proposed method until the norm of the gradient vector $\|\boldsymbol{\delta}\|$ converges to nearly 0 for different values of η. The results are shown in Figure 1, from which we can say that the conformity metrics can be learned quite efficiently if η is set to a value not too small. In this experiment, for η that is equal to or greater than 10, around 10 iterations and a few tens of iterations are enough for the Cosme and Anime review datasets, respectively.

We further investigated the distribution of the resulting conformity metrics and plotted them for each value of η in Figure 2. It is clear that the trends observed from these figures are quite similar for both the datasets. Namely, with adequate values of η, the majority of the users have the average value, i.e., 1.0 and only a small fraction of them have high or low conformity metric[5]. These characteristic users are worthy of attention. It is expected that the user with high conformity metric provides a consensus score early on[6]. Thus, such scores are worth considering. On the other hand, those who have low

[5] We did not place a constraint that the average of the conformity metric is 1.0, but the results are indeed very close to 1.0 for each η.

[6] Later, we see that this is not necessarily true.



(a) Cosme dataset (b) Anime dataset

Fig. 2. Distributions of the resulting conformity metrics.

conformity metrics tend to rate an item in a different way from the majorities. Thus, their scores may be useful for a particular user with similar interests. Further, knowing these low scores help preventing other users from being confused by such unusual ratings.

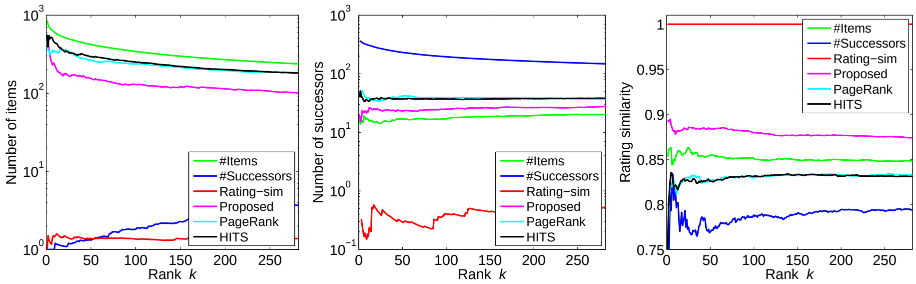
4.2 Evaluation of conformity metrics

In this section, we investigate how a user with high conformity metric can be characterized by other naive measurements. For this purpose, we focus on the following three properties which are naturally considered as the necessary conditions for a user to be high conformity metric: 1) she rates many items compared to other users (number of items), 2) she has many followers who rate the same items that she rated (number of successors), and 3) the scores of the followers are similar to those she rated (rating similarity). We first ranked users in each dataset according to i) these three naive measurements and ii) the conformity metrics returned by our proposed method, and investigated how top-K ranked users for each metric perform for other metrics, e.g., Do the top-K ranked users of "rating similarity" perform good or bad for "number of successors" and "number of items"? The number of successors of user v is given by $(1/|I(v)|)\sum_{i\in I(v)}|W(i,t_{v,i})|$, while the rating similarity between the score of user v and ones given by its followers is defined as follows:

$$1 - \frac{1}{|I(v)|}\sum_{i\in I(v)}\frac{|s_{v,i} - \frac{1}{|W(i,t_{v,i})|}\sum_{u\in W(i,t_{v,i})} s_{u,i}|}{|S|-1} \tag{4}$$

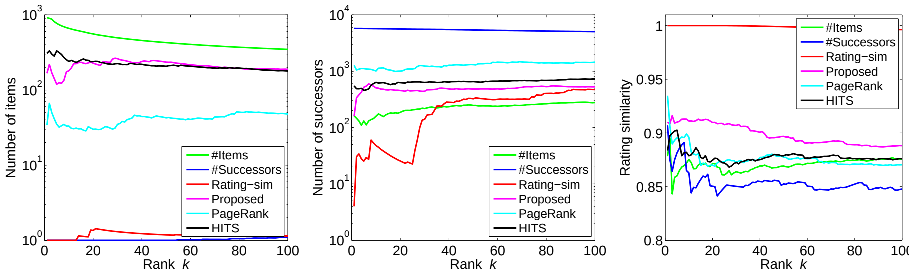
Hereafter, we refer to each ranking method as "#Items", "#Successors", "Rating-sim", and "Proposed", respectively according to the measurement metric used.

We, second, compared the results by the above ranking methods with those obtained by PageRank [1] and HITS [4]. These two algorithms rank nodes in a network. Tang J. et.al. (2013) [12] used a trust network and ranked nodes by PageRank and used this information to place a weight on each node to reflect conformity metric of each node.



(a) Number of items (b) Number of successors (c) Rating similarity

Fig. 3. Comparison of the top 282 users derived from the Cosme review dataset by using the six methods in the three measures.



(a) Number of items (b) Number of successors (c) Rating similarity

Fig. 4. Comparison of the top 100 users derived from the Anime review dataset by using the six methods in the three measures.

The trust network is a different source of information, i.e., social relations, and no such information is available to us for the two review systems we are using. Thus, we approximately induced a trust network from the observed score records by considering users as nodes and linking user u and v with a directed link (u, v) (link from u to v) if user v has rated an item i before user u rates it. This linking method shares, with our proposed method, the idea that a user rates an item considering scores already given to the same item by others. We set the teleportation probability of the PageRank algorithm to a typical value, i.e., 0.15, and used the authority score to rank nodes for the HITS algorithm. We call these ranking methods as "PageRank" and "HITS" in what follows.

The results for the Cosme and Anime review datasets are shown in Figures 3 and 4, respectively. Each line depicts the average score of the mentioned metric (see the title) for the corresponding ranking method in the caption box (one of #Items, #Successors, Rating-sim, Proposed, PageRank, and HITS) over the top k users. Thus, the mentioned method always gives the best result in Figures 3 and 4. In this experiment, we adopted $K = 100$. But, for the Cosme review dataset, we actually considered the top 282 users because rating similarity score is the same for the top 282 users, which is 1.0. We set η to 20 for the Cosme dataset, and 10 for the Anime dataset, respectively in the proposed method because they achieved the best (lowest) perplexity in the previous experiment.

In Figures 3(a) and 4(a), the number of items by Proposed is much larger than those by #Successors and Rating-sim. It is noted that, in an extreme case, even if user v rated only one item, v can achieve a high score in the number of successors if many users rate the item after she did. Similarly, her rating similarity can be large if her followers give the same score as she did even if the number of items she rated is only one. This is the reason why the resulting values by #Successors and Rating-sim tend to be quite small in these figures. This implies that users who make #Items, #Successors and Rating-sim large separately are different from each other.

In Figures 3(b) and 4(b), the results by Proposed achieves better scores than those by #Items and Rating-sim. Likewise, Proposed outperforms #Items and #Successors in terms of the rating similarity as in Figure 3(c) and 4(c). From these results, we can say that the conformity metrics learned by Proposed can be a good indicator to identify those who satisfy the above three necessary conditions simultaneously. In addition, from Figures 3(b) and 4(b), it is found that the number of successors of the users having the highest conformity metric is not necessarily large. This means a user with high conformity metric is not necessarily an early adopter.

On the other hand, it is suggested from Figures 3(a) and 4(a) that both PageRank and HITS can be better indicators than #Successors and Rating-sim to identify those users who rate many items. This is because these two ranking methods tend to give a higher score to a node that has a larger in-degree. Similarly, they outperform #Items and Rating-sim in terms of the number of successors in Figures 3(b) and 4(b). However, in Figure 3(c) and 4(c), their scores are comparable to or lower than those by #Items. This means that the scores of PageRank and HITS are no better indicators than the conformity metrics of Proposed to identify users who satisfy the three basic properties altogether.

5 Conclusion

In this paper, we addressed the problem of quantitatively assessing the conformity of a user in the context of rating items, and proposed an efficient algorithm that learns the conformity metric of each user from observed review scores. The idea behind is that a user often rates an item taking into account not only her own opinion but also scores already given to the item by other users, and the reliability of scores depend on who rated them. We modeled this rating process as a stochastic decision making process and used a modified Voter model. The proposed method can efficiently learn the conformity metrics based on an EM-like algorithm within a few tens of iterations. Its generalization capability is insensitive to the value of the regularization factor. Empirical evaluation on the two real world review datasets uncovered some interesting findings about the conformity metrics learned by the proposed algorithm. The majority of people have an average conformity metric with adequate regularization factors, i.e., 1.0 and only a limited fraction of people have high or low conformity metrics, who are worth paying attention to. Conformity metric can be a good indicator to identify those who satisfy the following three basic properties simultaneously that are considered natural for a user to be of high conformity, i.e., 1) a multitude of rated items, 2) a multitude of followers, and 3) a high rating similarity between her own scores and her follower's. None of them can be a good indicator alone. We further found that users having a high PageRank score or a high HITS score tend to rate a large number of items and have a large number of followers, satisfying the above two properties, but their rating similarity is not as large as that of those who have high conformity metrics or those who rate a large number of items.

Acknowledgments

This work was partly supported by Asian Office of Aerospace Research and Development, Air Force Office of Scientific Research under Grant No. AOARD-13-4042, and JSPS Grant-in-Aid for Scientific Research (C) (No. 26330352).

References

1. Brin, S., Page, L.: The anatomy of a large-scale hypertextual web search engine. Computer Networks and ISDN Systems 30, 107–117 (1998)
2. Even-Dar, E., Shapria, A.: A note on maximizing the spread of influence in social networks. In: Proceedings of the 3rd International Workshop on Internet and Network Economics (WINE'07). pp. 281–286 (2007)
3. Jamali, M., Ester, A.: A matrix factorization technique with trust propagation for recommendation in social networks. In: Proceedings of the 4th ACM Conference on Recommender Systems (RecSys'10). pp. 135–142 (2010)
4. Kleinberg, J.: Authoritative sources in a hyperlinked environment. Journal of the ACM 46(5), 604–632 (1999)
5. Koren, Y.: Factorization meets the neighborhood: A mulifaceted collaborative filtering model. In: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'08). pp. 426–434 (2008)
6. Ma, H., King, I., Lyu, M.R.: Learning to recommend with social trust ensemble. In: Proceedings of the 32nd International ACM SIGIR Conference on Research and Development (SIGIR'09). pp. 203–210 (2009)
7. Ma, H., Yang, H., Lyu, M.R., King, I.: Sorec: Social recommendation using probabilistic matrix factorization. In: Proceedings of the 17th ACM Conference on Information and Knowledge Management (CIKM'08). pp. 931–940 (2008)
8. Ma, H., Zhou, D., Liu, C., Lyu, M.R., King, I.: Recommender systems with social regularization. In: Proceedings of the 4th ACM International Conference on Web Search and Data Mining (WSDM'11). pp. 287–296 (2011)
9. Michael, L., Otterbacher, J.: Write like I write: Herding in the language of online reviews. In: Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media (ICWSM '14). pp. 356–365 (2014)
10. Salakhutdinov, R., Mnih, A.: Probabilistic matrix factorization. In: Advances in Neural Information Processing Systems 20 (NIPS'07). pp. 791–798 (2008)
11. Sood, V., Redner, S.: Voter model on heterogeneous graphs. Physical Review Letters 94, 178701 (2005)
12. Tang, J., Hu, X., Gao, H., Liu, H.: Exploiting local and global social context for recommendation. In: Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI'13). pp. 2712–2718 (2013)