Certainty-based Preference Completion Postprint

Authors: Lei, Li, Minghe, Xue, Zan, Zhang, Huanhuan Chen, Xindong, Wu, Lei, Li

Date: 2022-11-28T00:00:00+00:00

Abstract

Since it is often impractical to ask agents to provide linear orders over all alternatives, preference completion is necessary for these partial rankings. Specifically, the personalized preferences of each agent over all alternatives can be estimated using partial rankings from neighboring agents over subsets of alternatives. However, since agents' rankings are nondeterministic and may contain noise, it is both necessary and important to conduct certainty-based preference completion.

Hence, in this paper, we first construct a bijection from the ranking space to the preference space for alternative pairs with the obtained ranking set, and evaluate the certainty and conflict of alternative pairs using a well-established statistical measurement—the Probability-Certainty Density Function based on subjective probability. Then, a certainty-based voting algorithm that incorporates both certainty and conflict is employed to perform certainty-based preference completion. Moreover, the properties of the proposed certainty and conflict measures have been studied empirically, and the proposed approach for certainty-based preference completion of partial rankings has been experimentally validated against state-of-the-art approaches using several datasets.

Full Text

Preamble

RESEARCH PAPER

Certainty-based Preference Completion

Lei Li[1],[2]†, Minghe Xue[2], Zan Zhang[2], Huanhuan Chen[3] & Xindong Wu[1]

[1]Key Laboratory of Knowledge Engineering with Big Data (Hefei University of Technology), Ministry of Education, Hefei 230009, China

[2]School of Computer Science and Information Engineering, Hefei University of Technology, Hefei 230009, China
[3]School of Computer Science and Technology, University of Science and Technology of China, Hefei 230026, China

Keywords: Preference completion; Nondeterministic; Certainty; Subjective probability; Conflict

Citation: Li, L., et al.: Certainty-based preference completion. Data Intelligence 4(1), 112-133 (2022). doi: 10.1162/dint_a_{00115}

Received: October 15, 2021; **Revised:** December 7, 2021; **Accepted:** January 11, 2022

ABSTRACT

Since it is often impractical to require agents to provide complete linear orders over all alternatives, preference completion becomes necessary for handling partial rankings. Specifically, the personalized preferences of each agent over all alternatives can be estimated using partial rankings from neighboring agents over subsets of alternatives. However, because agents' rankings are nondeterministic and may contain noise, it is both necessary and important to conduct certainty-based preference completion.

This paper makes several contributions in this domain. First, for alternative pairs with an obtained ranking set, we establish a bijection from the ranking space to the preference space, and evaluate the certainty and conflict of alternative pairs using a well-constructed statistical measurement called the Probability-Certainty Density Function based on subjective probability. Then, we propose a certainty-based voting algorithm that leverages both certainty and conflict to perform preference completion. Furthermore, we empirically study the properties of the proposed certainty and conflict measures, and experimentally validate our approach against state-of-the-art methods using several datasets.

---

1. INTRODUCTION

In preference completion problems, we are given a set of agents (users) and a set of alternatives (items), where each agent provides a partial ranking over a subset of alternatives. The goal is to infer each agent' s personalized ranking or preference over all alternatives, including those not yet evaluated by the agent. Clearly, it is often impractical to ask agents to provide linear orders over all alternatives, particularly in big data environments [?]. For instance, an agent may be unaware of some alternatives due to their sheer number, making it difficult to rank all of them. Alternatively, some alternatives may be incomparable for a particular agent. These situations result in partial rankings, necessitating preference completion.

The preference completion problem finds applications in numerous domains, including social choice and recommender systems [?], where it proves useful for community detection [?, ?] and graph anomaly detection [?]. In social choice, for example, each voter (agent) can cast a ballot as either a complete ranking over all candidates (alternatives) or a partial ranking over some candidates. For these partial rankings, a voting rule must be applied to form a ranking over all candidates. In recommendation systems, each user can rate some items, and the system' s task is to predict ratings for unrated items. Two common approaches address this requirement: matrix factorization and neighborhood-based methods. While traditional algorithms for these approaches are typically rating-oriented, recent work has shifted focus to ranking-oriented algorithms [?, ?] due to limitations of rating-oriented methods. This paper concentrates on the ranking-oriented neighborhood-based approach.

Traditional neighborhood-based preference completion first identifies near neighbors for each agent and then aggregates these neighbors' rankings using a voting rule to produce predicted preferences [?]. However, this approach faces inevitable challenges. For example, agents may exhibit irrational behavior or provide rankings in noisy settings. To address these issues, many rating-oriented trust-based approaches have been proposed that incorporate additional contextual information, while ranking-oriented approaches offer substantial room for improvement. Liu et al. [?] proposed an anchor-based algorithm that leverages ranking information from many other agents to mitigate the effects of randomness.

Building upon Liu' s work [?], this paper proposes a certainty-based preference completion algorithm. Specifically, after identifying k-nearest neighbors using Liu' s anchor-kNN algorithm, we employ the certainty-based voting algorithm introduced herein to complete preferences (rankings) instead of using traditional majority voting. The majority voting rule is prone to erroneous judgments, especially when votes are closely split, as even slight randomness can alter outcomes. To address this, we introduce a certainty-based voting algorithm that incorporates a certainty measure—quantifying the degree to which two alternatives can be preferentially ordered or compared. Only when the certainty value meets a defined threshold do we proceed with a three-way preference decision rather than simply assigning a binary preference. Consequently, the certainty-based voting algorithm avoids incorrect judgments when scores are close or rankings are noisy.

Before formalizing certainty and presenting our algorithm, we first examine the relationship between certainty and preference space to introduce three-way preference between two alternatives. Technically, in a ranking pool gathered from agents, rankings containing alternative pair A and B can be aggregated to form the preference between A and B. Mathematically, we can construct a bijection from the ranking space to the preference space for alternative pair

A and B. The ranking space comprises all partial rankings on A and B from agents, while the preference space consists of three-way preference between A and B: preference (prefer A to B, denoted as AB_P), dispreference (prefer B to A, denoted as BA_P), and uncertainty (no preference between A and B, denoted as AB_C), following trisecting and acting models of human cognitive behaviors [**?**, **?**]. This framework distinguishes three situations: (1) agents prefer A to B, confirmed by high AB_P and low BA_P and AB_C; (2) agents prefer B to A, confirmed by low AB_P and high BA_P and low AB_C; and (3) agents are uncertain about the preference between A and B, i.e., A and B are incomparable, confirmed by low AB_P, low BA_P, and high AB_C.

When AB_C is low, the preference between A and B can be determined, meaning A and B are comparable. Therefore, we introduce the certainty of preference, denoted as C, to describe the trustworthiness of the preference. Following the proposition that certainty represents the degree of belief an individual has in a preference [**?**], we can treat certainty as the subjective probability of the preference. In this paper, we evaluate certainty using a well-constructed statistical measurement that defines a bijection from ranking space to preference space, enabling estimation of pairwise preferences from neighbors' partial rankings via mapping to AB_C. The certainty can be calculated as $C = 1 - \int |f(O|X) - f(X|O)| dX$, where $f(O|X)$ is the posterior distribution and $f(X|O)$ is the prior (uniform) distribution.

Our certainty definition should capture two key properties: (1) certainty increases as the number of rankings between alternative pair A and B increases for a fixed ratio of rankings preferring A to B versus B to A; and (2) certainty decreases as the extent of conflict increases in the partial rankings between A and B.

Our main contributions are: (1) we introduce probability-based certainty and conflict measures under Properties 1 & 2 to describe preference trustworthiness, as suggested in [**?**] that these may be more important than preference itself; (2) we propose a certainty-based voting algorithm using certainty and conflict for preference completion in nondeterministic settings; and (3) we empirically study our approach' s properties and experimentally validate it against state-of-the-art methods on several datasets.

This paper is organized as follows. Section 2 reviews the Plackett-Luce model, Kendall-Tau distance, and anchor-kNN algorithm. Section 3 establishes a bijection from ranking space to preference space and evaluates certainty and conflict using the Probability-Certainty Density Function. Section 4 presents the certainty-based voting algorithm for preference completion. Section 5 empirically studies the properties of certainty and conflict. Section 6 experimentally validates our approach against state-of-the-art methods. Finally, Section 7 concludes and outlines future work.

2.1 Plackett-Luce Model

Given a set of m alternatives and a set of n agents, let $\mathbf{y} = (y_1, y_2, \ldots, y_m)$ denote the latent features of alternatives and $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ denote the latent features of agents. Agent i' s ranking R_i is determined by a statistical model for ranking data. As a widely-used statistical model, the Plackett-Luce model [**?**, **?**] is adopted to generate agent rankings. In this paper, each alternative is assigned a positive utility value; the greater this utility, the more likely its corresponding alternative is ranked at a higher position [**?**]. Following [**?**], the realized utility for every alternative j on agent i is determined by $u_{ij}(x_i, y_j) = h(x_i, y_j) + e_{i,j}$, where $h(x_i, y_j)$ is agent i' s expected utility on alternative j and can be determined by the closeness of latent features x_i and y_j, measured by $h(x_i, y_j) = \exp(-\|x_i - y_j\|^2)$, and $e_{i,j}$ is a zero-mean independent random variable following a Gumbel distribution. When the realized utilities set $\mathbf{u}_i = (u_{i1}, u_{i2}, \ldots, u_{im})$ of agent i is obtained, agent i ranks alternatives in decreasing order according to realized utilities. Repeating this process n times generates synthetic datasets for all agents. For more details, please refer to Algorithm 1.

Algorithm 1. Sampling from Plackett-Luce Model

2.2 Kendall-Tau Distance

Given two agents' rankings R_1 and R_2 over the same alternatives, the Kendall-Tau distance measures their similarity as the total number of disagreements in pairwise comparisons between alternatives in the linear rankings. For alternative j in R_i, $R_i(j)$ represents its position in R_i. For example, if j is the top-ranked alternative in R_i, then $R_i(j) = 1$. The normalized Kendall-Tau distance between R_1 and R_2 is:

$$NK(R_1, R_2) = \frac{\sum_{j,k} I(R_1(j) < R_1(k) \wedge R_2(j) > R_2(k))}{\binom{|R_1 \cap R_2|}{2}}$$

where $I(v)$ is an indicator function set to 1 if argument v is true and 0 otherwise. Moreover, if rankings do not share exactly the same alternatives, the intersection of the two alternative sets can be used to compute the normalized Kendall-Tau distance.

2.3 Anchor-kNN Algorithm

Before introducing the anchor-kNN algorithm proposed in [**?**], we first present KT-kNN, which simply uses Kendall-Tau distance to find an agent' s neighbors. If the Kendall-Tau distance between two rankings R_i and R_j is small, the latent features of agents x_i and x_j should be close, indicating similar opinions on alternatives.

Since KT-kNN does not consider that agents' preferences may be nondeterministic or that rankings may be made in noisy settings, anchor-kNN differs by using other agents' ranking data (named anchors) to determine the closeness between two agents rather than considering only the two agents' rankings. Anchor-kNN develops a feature $F_{i,j}$ for agents i and j representing the Kendall-Tau distance between R_i and R_j, i.e., $F_{i,j} = NK(R_i, R_j)$. To measure the closeness between two agents denoted as $D_{i,j}$, we use the sum of differences between $F_{i,t}$ and $F_{j,t}$ to find k-nearest neighbors, where t is a third agent belonging to all other agents except i and j.

---

3. CERTAINTY AND PREFERENCE SPACE

This section presents preliminary definitions. For an arbitrary alternative pair A and B, certainty can describe the trustworthiness of the preference between them. Following [**?**], a Probability-Certainty Density Function (PCDF) can capture the subjective probability of rankings. However, unlike [**?**] and following [**?**, **?**], this paper defines certainty based on PCDF to satisfy Properties 1 & 2.

3.1 Ranking Space

The ranking space consists of all weighted partial rankings on alternative pair A and B from agents, including: rankings AB_O where A is ranked ahead of B with weight w_{AB}; rankings BA_O where B is ranked ahead of A with weight w_{BA}; and unordered rankings AB_O where A and B are not comparable with weight w_{AB}. Let n_{AB} denote the accumulated weight of rankings AB_O, n_{BA} denote the accumulated weight of rankings BA_O, and n_{AB} denote the accumulated weight of rankings AB_O. Obviously, we have $n_{AB} + n_{BA} + n_{AB} = 1$. Moreover, the weight w_{AB} for AB_O represents the quality of ranking AB_O. Without additional knowledge, we assign $w_{AB} = 1$.

DEFINITION 1. Ranking space $\mathcal{R} = \{\langle n_{AB}, n_{BA}, n_{AB}\rangle \mid n_{AB}, n_{BA}, n_{AB} \geq 0, n_{AB} + n_{BA} + n_{AB} = 1\}$.

3.2 Preference Space

Traditionally, uncertainty is often ignored, and sometimes dispreference is not considered either, leading to disturbing results as shown in the empirical study section. According to trisecting and acting models of human cognitive behaviors [**?**, **?**], the preference space consists of three-way preference between alternatives: preference AB_P (prefer A to B), dispreference BA_P (prefer B to A), and uncertainty AB_C (no preference between A and B).

DEFINITION 2. Preference space $\mathcal{P} = \{\langle AB_P, BA_P, AB_C \rangle \mid AB_P, BA_P, AB_C \in [0,1], \min\{AB_P, BA_P, AB_C\} > 0\}$.

3.3 Certainty of Rankings in Alternative Pairs

Bayesian inference [**?**, **?**] is adopted here to update probabilities with available contextual information about rankings in alternative pairs, i.e., to update the prior distribution to the posterior distribution [**?**, **?**]. This paper utilizes offline Bayesian inference, though it can also be applied to online/streaming scenarios [**?**, **?**].

Let x_{AB}, x_{BA}, and x_{AB} be the probabilities of rankings AB_O, BA_O, and AB_O, respectively, where $x_{AB}, x_{BA}, x_{AB} \in [0,1]$ and $X = \langle x_{AB}, x_{BA}, x_{AB} \rangle$. Thus, we have $x_{AB} + x_{BA} + x_{AB} = 1$.

Without additional information, the prior distribution $f(X|O)$ is uniform. Since the cumulative probability of a distribution within $[0,1]^2$ equals 1, the density of a PCDF has mean value 1 within $[0,1]^2$, making $f(X|O) = 1$.

As the ranking sample O conforms to a multinomial distribution [**?**, **?**], we have $f(O|X) \propto x_{AB}^{n_{AB}} x_{BA}^{n_{BA}} x_{AB}^{n_{AB}}$. For the posterior distribution $f(X|O)$, it can be estimated as [**?**, **?**]: $f(X|O) = \frac{f(O|X)f(X|O)}{\int f(O|X)f(X|O)dX}$.

Certainty can then be determined by the deviations of the posterior distribution from the prior (uniform) distribution. Hence, we have the following definition:

DEFINITION 3. The certainty AB_C of rankings $\langle n_{AB}, n_{BA}, n_{AB} \rangle$ can be estimated as $AB_C = 1 - \frac{1}{2}\int |f(X|O) - f(X|O)|dX$, where the factor $\frac{1}{2}$ removes double counting of deviations.

From this definition, we have $AB_C \in [0,1]$.

3.4 Conflict of Rankings in Alternative Pairs

Conflict is determined by the relative difference between weighted rankings n_{AB} and n_{BA}, as in [**?**]. Specifically, conflict is maximal when $n_{AB} = n_{BA}$ and minimal when $n_{AB} = 0$ or $n_{BA} = 0$.

DEFINITION 4. The conflict c_{AB} of rankings $\langle n_{AB}, n_{BA}, n_{AB} \rangle$ can be estimated as $c_{AB} = \frac{\min\{n_{AB}, n_{BA}\}}{n_{AB} + n_{BA}}$.

From this definition, we have $c_{AB} = c_{BA}$.

3.5 Bijection from Ranking Space to Preference Space

With Definitions 1, 2, 3, and 4, we introduce the following mapping:

DEFINITION 5. The bijection from ranking space $\mathcal{R} = \{\langle n_{AB}, n_{BA}, n_{AB} \rangle\}$ to preference space $\mathcal{P}$ can be estimated as $AB_P = n_{AB} \cdot AB_C$, $BA_P = n_{BA} \cdot AB_C$, and $AB_C = 1 - AB_P - BA_P$.

4. CERTAINTY-BASED PREFERENCE COMPLETION

This section proposes our certainty-based preference completion approach, with the framework shown in Figure 1 [Figure 1: see original paper]. The approach includes two processes: (1) finding k-nearest neighbors for user i using Liu' s anchor-kNN algorithm [?], and (2) conducting linear ranking for user i over all alternatives. We focus on the latter process, which uses neighbors' partial rankings and our certainty-based voting algorithm to estimate pairwise preferences for all alternative pairs, subsequently forming a linear ranking for user i.

4.1 Certainty-based Voting Algorithm

First, we introduce a key definition. **DEFINITION 6.** With preference space $\mathcal{P} = \{\langle AB_P, BA_P, AB_C \rangle\}$, the following conclusions obtain: if $AB_C \geq e_1$, alternatives A and B are unpreferred; if $AB_P - BA_P \geq e_2$, user i prefers A to B; if $BA_P - AB_P \geq e_2$, user i prefers B to A; otherwise, A and B are unpreferred, where e_1 and e_2 are thresholds to rule out comparison fuzziness.

Existing work using k-nearest neighbors typically applies common voting rules (e.g., majority voting) to estimate pairwise preferences for preference completion. Common voting rules may include positional scoring rules, maximin, and Bucklin; see [?] for details.

In contrast, this paper employs a certainty-based voting rule using certainty and conflict to obtain pairwise preferences. Certainty and conflict measure the trustworthiness that pair alternatives can be preferred or compared. If certainty satisfies a defined threshold, we evaluate the degree to which user i prefers one alternative over another, denoted by AB_P. Only when the difference between two-way preference reaches a certain value do we make a preference decision. Technically, for alternative pair A and B with $AB_P - BA_P \geq e_2$, a preference decision can be made. Algorithm 2 shows the process for estimating pairwise preference. Applying this algorithm to all alternative pairs yields all pairwise preferences.

Algorithm 2. Certainty-based voting algorithm for estimating pairwise preference

4.2 Greedy Order Algorithm

Next, we combine all pairwise preferences to form a linear ranking over all alternatives. One effective approach is the greedy order algorithm [?]. This algorithm follows a greedy strategy: it always picks the alternative with maximum potential value from the alternatives pool $\mathcal{I}$ and ranks it above all remaining items. For item i, the potential value v_i equals $\sum_{j \in \mathcal{I}} AB_P$. This value aggregates all pairwise preferences from the previous step and represents the preference for item i among all neighbors' rankings. The algorithm then deletes the selected item from the pool and updates potential values of remaining items by removing the selected item' s effects. Repeating this process until the alternatives pool is empty produces a linear ranking for user i. See Algorithm 3.

Algorithm 3. Greedy order algorithm

---

5. EMPIRICAL STUDIES ON PROPERTIES OF CERTAINTY

This section examines the properties of certainty and conflict in our proposed model.

5.1 Increasing Rankings with Fixed Conflict

Figure 2 [Figure 2: see original paper] plots how certainty AB_C varies with weighted rankings n_{AB} and n_{BA} under fixed conflict c_{AB}, which should confirm Property 1.

THEOREM 1. For fixed c_{AB} and $n_{AB} + n_{BA}$, certainty AB_C increases with $n_{AB} + n_{BA}$.

Proof: Let $a = n_{AB} + n_{BA} = b$, and $f(x) = x_{AB}^{n_{AB}} x_{BA}^{n_{BA}} (1 - x_{AB} - x_{BA})^{n_{AB}}$. As in [**?**], we can define x_1, x_2, x_3, x_4 such that $f(x_1) = f(x_2) = f(x_3) = f(x_4) = 1$ and $|f(x) - 1| \geq 0$. Then:

$$AB_C = 1 - \frac{1}{2} \int |f(x) - 1| dx = 1 - \frac{1}{2} \left(\int_{x_1}^{x_2} (1 - f(x)) dx + \int_{x_3}^{x_4} (1 - f(x)) dx \right)$$

where x_1, x_2, x_3, x_4 are functions of b. Following Lemma 9 in [**?**], we have $\frac{\partial AB_C}{\partial b} > 0$, confirming Theorem 1.

5.2 Increasing Conflict with Fixed Rankings

Figure 3 [Figure 3: see original paper] plots how certainty AB_C varies with weighted rankings n_{AB} and n_{BA} under fixed $n_{AB} + n_{BA}$ and fixed n_{AB}, which should confirm Property 2.

THEOREM 2. For fixed $n_{AB} + n_{BA}$, certainty AB_C is decreasing with $n_{AB} \leq n_{BA}$ and increasing with $n_{AB} \geq n_{BA}$.

Proof: The validation details are omitted as they are similar to Theorem 1' s proof. Specifically, by removing the absolute value sign and differentiating, we can prove the derivative is negative for $n_{AB} \leq n_{BA}$ and positive for $n_{AB} \geq n_{BA}$.

---

6. EXPERIMENTS

This section examines the empirical performance of our certainty-based preference completion algorithm, comparing it with common majority voting [?] and classic collaborative filtering (CF) [?]. Both our algorithm and majority voting use anchor-kNN to find k-nearest neighbors' rankings for preference completion, while CF is a rating-oriented approach that computes user similarity to find neighbors and uses their ratings for prediction.

6.1 Datasets

Experiments use two dataset types. The synthetic dataset is created by sampling from a Plackett-Luce model using Algorithm 1, containing over 20,000 rankings from agents on 20 alternatives, each following a Gumbel distribution. The Flixster dataset collects movie ratings with social trust, containing over 8,000,000 ratings on over 2,000 movies. For experiments, we convert ratings to rankings and select over 9,000 rankings on over 50 movies.

6.2 Evaluation Metrics

We evaluate performance using three metrics: (a) prediction error, (b) Spearman correlation coefficient, and (c) Kendall rank correlation coefficient. The first measures predicted ranking quality, while the others measure correlation with the original ranking. See [?] and Liu et al. [?] for details.

Evaluation Metric 1: Prediction error estimates accuracy between predicted and true rankings:

$$\text{Prediction Error} = \frac{1}{M} \sum_{i,j,k} |Y_{i,j,k} - X_{i,j,k}|$$

where M is the maximum pairwise error, $Y_{i,j,k} = 1$ means user i prefers alternative j to k in the predicted ranking, $X_{i,j,k} = 1$ represents the same preference in the original ranking, and $I^{-}(v)$ equals 1 when $v < 0$ and 0 otherwise.

Evaluation Metric 2: Spearman correlation coefficient measures position differences for every alternative between predicted and original rankings:

$$\text{Spearman CC} = 1 - \frac{6 \sum_i d_i^2}{N(N^2 - 1)}$$

where d_i is the position difference for alternative i between predicted and original rankings.

Evaluation Metric 3: Kendall rank correlation coefficient uses Kendall distance to measure correlation:

$$\text{Kendall CC} = \frac{\sum_{j,k} I(Y_{i,j,k} = X_{i,j,k})}{\binom{|I_x \cap I_y|}{2}}$$

where symbols match Evaluation Metric 1, I_x is the alternative set in the original ranking, and I_y is the set in the predicted ranking.

6.3 Experimental Results on Synthetic Dataset and Flixster Dataset

This section presents comparison results using the evaluation metrics.

(a) Synthetic dataset. Figure 4 [Figure 4: see original paper] shows prediction error is smaller with our certainty-based algorithm than with CF and majority voting. Ranking-oriented approaches outperform the rating-oriented approach because rankings contain more preference relation information than rating scores, making neighbor identification and preference completion easier and more accurate. The comparison between certainty-based and majority voting demonstrates that considering certainty reduces randomness effects.

Figure 5 Figure 5: see original paper shows Spearman correlation coefficient performance, where our certainty-based algorithm outperforms others. By considering preference space and certainty, our approach filters out pair preferences with close votes and low certainty, making predicted rankings more trustworthy. Figure 5(b) shows similar results for Kendall rank correlation coefficient.

(b) Flixster dataset. Figure 6 [Figure 6: see original paper] shows that when $k > 300$, our approach outperforms others, with ranking-oriented methods still superior to rating-oriented methods. When $k < 300$, results are less optimal, possibly because converting rating data to ranking data introduces pairwise preference errors. With more neighbors, our algorithm demonstrates superiority, and prediction error decreases as neighbor count grows.

Figure 7 Figure 7: see original paper shows our certainty-based approach significantly outperforms others, consistent with synthetic dataset results. Figure 7(b) shows similar performance patterns.

Overall, experiments on both datasets validate our proposed certainty-based preference completion algorithm.

---

7. CONCLUSION AND FUTURE WORK

Since agents'rankings are nondeterministic and may be provided under noisy conditions, certainty-based preference completion is necessary and important. This paper first built a bijection from ranking space to preference space for alternative pairs and evaluated certainty and conflict using the Probability-Certainty Density Function. Then, a certainty-based voting algorithm using certainty and conflict conducted preference completion. Rankings with high certainty and low conflict can be obtained to complete preferences. We empirically studied the properties of our approach and experimentally validated it against state-of-the-art methods on several datasets.

In real applications, data is usually unbalanced [**?**], with some alternative pairs having many rankings while others have few. Future work will propose algorithms to handle unbalanced preference completion effectively and efficiently.

ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (No. 62076087, No. 61906059 & No. 62120106008) and the Program for Changjiang Scholars and Innovative Research Team in University (PCSIRT) of the Ministry of Education of China under grant IRT17R32.

The first author thanks his wife Jun Zhang, his parents, and friends during his fight with lung adenocarcinoma: "I leave no trace of wings in the air, but I am glad I have had my flight."

AUTHOR CONTRIBUTIONS

All authors including L. Li (lilei@hfut.edu.cn), M.H. Xue (18856337539@163.com), Z. Zhang (zanzhang@hfut.edu.cn), H.H. Chen (hchen@ustc.edu.cn), and X.D. Wu (xwu@hfut.edu.cn) participated in writing the paper. L. Li designed the algorithm and experiments and provided funding; M.H. Xue designed and conducted experiments and analyzed data; Z. Zhang analyzed data.

REFERENCES

[1] Li, L., et al.: Weighted partial order oriented three-way decisions under score-based common voting rules. International Journal of Approximate Reasoning 123(2020), 41–54 (2020)

[2] Liu, T.: Learning to rank for information retrieval. Berlin, Springer (2011)

[3] Liu, F., et al.: Deep learning for community detection: Progress, challenges and opportunities. In: Proceedings of the 29th International Joint Conference on Artificial Intelligence (IJCAI 2020), pp. 4981–4987 (2020)

[4] Su, X., et al.: A comprehensive survey on community detection with deep learning. arXiv preprint arXiv:2105.12584 (2021)

[5] Ma, X., et al.: A comprehensive survey on graph anomaly detection with deep learning. arXiv preprint arXiv:2106.07178 (2021)

[6] Katz-Samuels, J., Scott, C.: Nonparametric preference completion. In: Proceedings of International Conference on Artificial Intelligence and Statistics (AISTATS 2018), pp. 632-641 (2018)

[7] Liu, N., Yang, Q.: Eigenrank: A ranking-oriented approach to collaborative filtering. In: Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 83-90 (2008)

[8] Liu, A., et al.: Near-neighbor methods in random preference completion. In: Proceedings of the AAAI Conference on Artificial Intelligence (AAAI 2019), pp. 4336-4343 (2019)

[9] Yao, J.: Three-way granular computing, rough sets, and formal concept analysis. International Journal of Approximate Reasoning 116, 106-125 (2020)

[10] Li, L., Wang, Y.: Context based trust normalization in service-oriented environments. In: Proceedings of the IEEE Conference on Autonomic and Trusted Computing, pp. 122-138 (2010)

[11] Hallinan, J.T.: Why we make mistakes. Broadway Books, Portland (2010)

[12] Luce, R.: Individual choice behavior: A theoretical analysis. Dover Publications, New York (1959)

[13] Plackett, R.: The analysis of permutations. Applied Statistics 24, 193-202 (1975)

[14] Liu, A., et al.: Learning plackett-luce mixtures from partial preferences. In: Proceedings of the AAAI Conference on Artificial Intelligence (AAAI 2019), pp. 4328-4335 (2019)

[15] Jøsang, A.: A subjective metric of authentication. In: Proceedings of the 5th European Symposium on Research in Computer Security (ESORICS 98), pp. 329-344 (1998)

[16] Li, L., Wang, Y.: Subjective trust inference in composite services. In: Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence (AAAI 2010), pp. 1377-1384 (2010)

[17] Wang Y., Singh, M.P.: Evidence-based trust: A mathematical model geared for multiagent systems. ACM Transactions on Autonomous and Adaptive Systems 5(4), Article No. 14 (2010)

[18] Yao, Y.: Three-way decision: An interpretation of rules in rough set theory. In: Proceedings of the 4th International Conference on Rough Sets and Knowledge Technology (RSKT 2009), pp. 642-649 (2009)

[19] Chen, H., Tiño, P., Yao, X.: Probabilistic classification vector machines. IEEE Transactions on Neural Networks 20(6), 901-914 (2009)

[20] Chen, H., Tiño, P., Yao, X.: Predictive ensemble pruning by expectation propagation. IEEE Transactions Knowledge Data Engineering 21(7), 999-1013

(2009)

[21] Hamada, M.S., et al.: Bayesian reliability. Springer, Berlin (2008)

[22] Hines, W.W., et al.: Probability and statistics in engineering. John Wiley & Sons, Hoboken (2003)

[23] Chen, H., Tiño, P., Yao, X.: Efficient probabilistic classification vector machine with incremental basis function selection. IEEE Transactions on Neural Networks and Learning Systems 25(2), 356–369 (2014)

[24] Jiang, B., et al.: Scalable graph-based semi-supervised learning through sparse bayesian model. IEEE Transactions on Knowledge and Data Engineering 29(12), 2758–2771 (2017)

[25] Cohen, W.W., Schapire, R.E., Singer, Y.: Learning to order things. Journal of Artificial Intelligence Research 5, 243–270 (1999)

[26] Goldberg, D., et al.: Using collaborative filtering to weave an information tapestry. Communications of the ACM 35(12), 61–70 (1992)

[27] Fieller, E., Herman, H., Pearson, E.: Tests for rank correlation coefficients. I. Biometrika, 44(3/4), 470–481 (1957)

[28] Gong, Z., Chen, H.: Model-based oversampling for imbalanced sequence classification. In: Proceedings of the 25th ACM International Conference on Information and Knowledge Management (CIKM' 16), pp. 1009–1018 (2016)

AUTHOR BIOGRAPHY

Lei Li received his Bachelor' s degree from Jilin University, Changchun, China, in 2004, his Master' s degree from Memorial University of Newfoundland, St. John' s, Canada, in 2006, and his Ph.D. from Macquarie University, Sydney, Australia, in 2012. He is currently an Associate Professor of computer science and technology at the Key Laboratory of Knowledge Engineering with Big Data (Hefei University of Technology), Ministry of Education, China, Intelligent Interconnected Systems Laboratory of Anhui Province (Hefei University of Technology), and School of Computer Science and Information Engineering, Hefei University of Technology. His research interests include data mining, social computing, and graph computing. He is a senior member of IEEE. ORCID: 0000-0002-5374-7293

Minghe Xue received the BE degree from Hefei University of Technology, China, in 2019. Her research interests are in graph computing and social computing. She is currently a Master' s candidate in the Key Laboratory of Knowledge Engineering with Big Data (Hefei University of Technology), Ministry of Education, China, and School of Computer Science and Information Engineering at Hefei University of Technology. ORCID: 0000-0001-8016-5136

Zan Zhang received his Ph.D. in Computer Science from Hefei University of Technology, China, in 2018. He is currently a lecturer at Hefei University of Technology. His research interests include data mining and knowledge engineering. ORCID: 0000-0002-6383-1683

Huanhuan Chen received the B.Sc. degree from the University of Science and Technology of China (USTC), Hefei, China, in 2004, and the Ph.D. in computer science from the University of Birmingham, Birmingham, U.K., in 2008. He is currently a Full Professor with the School of Computer Science and Technology, USTC. His research interests include neural networks, Bayesian inference, and evolutionary computation. He received the 2015 International Neural Network Society Young Investigator Award, the 2012 IEEE Computational Intelligence Society Outstanding Ph.D. Dissertation Award, the IEEE Transactions on Neural Networks Outstanding Paper Award (bestowed in 2011 for a 2009 paper), and the 2009 British Computer Society Distinguished Dissertations Award. He is currently an Associate Editor for IEEE Transactions on Neural Networks and Learning Systems and IEEE Transactions on Emerging Topics in Computational Intelligence. He is a senior member of IEEE. ORCID: 0000-0002-3918-384X

Xindong Wu is a foreign academician of the Russian Academy of Science and Professor with the Key Laboratory of Knowledge Engineering with Big Data (Hefei University of Technology), Ministry of Education, China. He received his Bachelor’ s and Master’ s degrees in Computer Science from Hefei University of Technology, China, and his Ph.D. in Artificial Intelligence from the University of Edinburgh, United Kingdom. His research interests include data mining, knowledge engineering, and Web information exploration. Dr. Wu is a Fellow of IEEE and AAAS. He is the Steering Committee Chair of the IEEE International Conference on Data Mining (ICDM), Editor-in-Chief of Knowledge and Information Systems (KAIS, by Springer), and Editor-in-Chief of the Springer Book Series on Advanced Information and Knowledge Processing (AI and KP). He was Editor-in-Chief of IEEE Transactions on Knowledge and Data Engineering (TKDE) from 2005 to 2008. He served as program committee chair/co-chair for ICDM 2003, KDD 2007, CIKM 2010, and ICBK 2017. ORCID: 0000-0003-2396-1704