An Intransitivity Model for Matchup and Pairwise Comparison

Yan Gu
Graduate School of Informatics
Kyoto University

Jiuding Duan
Allianz Global Investors

Hisashi Kashima
Graduate School of Informatics
Kyoto University

***Abstract*—Ranking is a ubiquitous problem appearing in many real-world applications. The superior players or objects are oftentimes determined by a matchup or pairwise comparison. Various models have been developed to integrate the matchup results into a single ranking list of players and to further predict the results of future matchups. Amongst them, the Bradley-Terry model is a mainstream model that achieves the goals by constructing explicit probabilistic interpretation. However, the model suffers from its strong assumption of transitive relationships and becomes vulnerable in practices where intransitive relationships exist. Blade-Chest model is an alternative solution to this intransitivity challenge by allowing the multi-dimensional representation of players. In this paper, we propose a low-rank matrix approach to characterize all players and generalize the related works by introducing a unified framework. Our experimental results on synthetic datasets and real-world datasets show that the proposed model is stably competitive with the standard models in terms of the consistency of probabilistic model interpretation and the predictive performance in out-of-sample tests.**

I. Introduction

The modeling of pairwise comparisons and the modeling of ranking lists are used in a wide range of applications and share a common basis of interactions between players and objects. Representative pairwise comparisons are matchups, including online video games [1], [2] and sports tournaments [3], where the model of matchups are used to either pair an equal and fair game or to predict the winner by systematically taking into account of overall track records. In ranking problems where rankings are aggregated from partially observed results, models of comparisons and pairwise preferences are easily found as a founding element [4], [5] and the application includes recommender systems [6] and social choice systems [7]–[10].

The advantage of highlighting the pairwise comparisons for ranking problem comes in two. First, pairwise comparison is the basic element for the data-driven model and the learning system to read and to learn from, in order to predict a win/lose result of future competitions. Second, the interpretation of the model can be streamlined in the lens of a simple matchup matrix because the matchup matrix explicitly carves out the predictive ranking list out of the model definition and model parameters.

Two of the most widely used models of ranking from pairwise comparisons are the Thurstone model [9] and the Bradley-Terry (BT) model [11], [12]. Both of these two models belong to an extended class of Random Utility Models and share a common characteristic that a player in game is parameterized by a scalar score that represents his overall capability. In more recent related works [13], the BT model becomes the basis of this probabilistic approach and shows a well-balanced status between interpretability and predictability.

A critical limitation of the BT pairwise comparison model is that the strength or the competitiveness of a player is modeled by using only a single scalar. In fact, it is not sufficient to represent one player with only a single scalar because the general features of the matchups between players will remain transitive. It is straightforward to show that if player A beats B and B beats C more often than not, then the model will predict that A always beats C. However, intransitive relationship naturally exists in many real world applications such as economics, sports games and social choice theory. An exemplar case is a basketball tournament with three teams, i.e. team A, team B and team C. The transitive model can not satisfy the following potential outcomes at the same time: Pr(A beats B) > 0.5, Pr(B beats C) > 0.5 and Pr(C beats A) > 0.5, where Pr(A beats B) > 0.5 implies the probability of team A beats team B is larger than 0.5, that is, A beats B more often than not. Besides, certain essential factors, like injuries, teamwork and psychological factors, are also not explicitly included. As a result, the marginal impact of these factors is unknown. Therefore, these model limitations motivate solutions in such direction in the pursuit of edging improvement of the fairness in pairing players and the accuracy in predictive setting.

To overcome such limitation, several models of intransitive comparisons have been proposed, such as the 2-dimensional vector representation BT model [14], the Blade-Chest (BC) model [15], [16] that introduced two multi-dimensional vectors, i.e. 'blade' and 'chest', to reflect different aspects of each player's strength, and a generalized intransitive model [17] which carves the relationship between players out of the intrinsic characteristics of players. All of these works considered multi-dimensional representation of each player and used the stochastic gradient method to solve the optimization problem.

In this paper, we propose a low-rank matrix approach to develop an intransitive model which is generalized from the Blade-Chest model. The generalized model avoids the estimation of the 'blade' and 'chest' vectors of each player by introducing a low-rank relation matrix for each pairwise comparison. We show that the developed framework is efficient for intransitivity modeling, unifies a series of existing related works and constructively maintains an elastic link to deep neural network approaches. Experimental results show that the proposed model consistently outperforms the benchmark methods and excels the competitive deep learning approach by its simplicity.

II. Related work

The traditional and well-known models, Thurstone [9] and Bradley-Terry-Luce [11], [12], are the fundamentals of the matchup and pairwise comparison. They use one-dimensional embedding which assume that each item is completely represented by an inherent strength. These work were surveyed extensively [13], [18]. In addition, following the using of a single scalar to measure the strength of the players, the research for ranking of the players in real-world has also been widely studied. For instance, the matchmaking for online games [19], [20], sports [21]–[23] and rating system [24]–[26].

However, it is oversimple to use a single value to present the ability of each player in many real-world applications. It is unable to explore the intransitivity in the games like the rock-paper-scissor. [14] use a 2-dimensional vector to model the property of each player, but only with the verification on very small datasets. [27] uses a matrix factorization to predict scores of professional basketball games by using different feature for offense and defense, although it does not explicitly study the intransitive property . [15]–[17] propose state-of-the-art intransitive models that with multi-dimensional vector to represent each play which are analogous to the model [27]. The idea of multi-dimensional representation are appeared in [13], [28], but without intransitivity issues. Moreover, this idea has also been widely explored in many applications such as the recommendation system [27], [29], language modeling [30] and so on.

There are further extension of the multi-dimensional modeling to context-aware settings [15], which concerns learning from user's pairwise preferences to predict the choice and its probability and with a significantly improvement. Learning with context has also been studied in recommendation systems in which using context information can improve the performance [31]–[33], and context-aware decision-making where different contexts can change the decision [34], [35].

III. Problem setting and preliminaries

We start with some notations and define the problem of matchup and pairwise comparison modeling, and then review some preliminary models.

A. Notations and problem setting

In this work we focus on modeling comparisons between two players and we assume the result of each matchup cannot be a draw. Given a set of players P with $|P| = N$ and a dataset $\mathcal{D}$ which contains n pairwise matchup records (i, j), where $i, j \in P$. For any players $i, j \in P$, we denote $i \succ j$ when player i wins a match against player j. For an observed matchup between each pair of players can be described in 4-tuple as $(i, j, 1, 0)$ that means $i \succ j$, or $(i, j, 0, 1)$ which means $j \succ i$. In any subset of $\mathcal{D}$, the game between the same players can be aggregated, and also resulting in a 4-tuple as (i, j, n_i, n_j). In this data entry, n_i indicates the number of times i wins j and n_j means the opposition.

By learning the representation or the ability of the players from a training set in which contains multiple matches, we want to predict the result of any future matchups. In the following subsections, we review several models for modeling matchup or pairwise comparison based on the notion of the matchup matrix that gives the winning probability of a match between two players. Bradley-Terry (BT) model is the most basic model that only allows transitive ordering of the items. The Blade-Chest (BC) model allows intransitive ordering by generalizing the BT model. The Blade-Chest-Sigma model is a generalized intransitive model of the BC model. We will generalize the models in the next subsections.

B. Bradley-Terry model

Bradley-Terry model [11], [12] is the most basic model for pairwise comparison. A critical determinant of the Bradley-Terry model is matchup matrix $\mathbf{M} \in \mathcal{R}^{N\times N}$, whose entry M_{ij} indicates the comparative advantages of item i over item j. $M_{ij} > 0$ literally reads, "item i has a comparative advantage over item j." and vice versa.

In the model, each player i is given a strength parameter γ_i, and the (i, j)-th element of the matchup matrix can be defined as

$$M_{ij} = \gamma_i - \gamma_j.$$

When two players i and j play a match, the winning probability $p_{ij} = \Pr(i \succ j)$ that player i wins the match is given using the matchup matrix, that is,

$$\begin{aligned} p_{ij} &= \frac{\exp(\gamma_i)}{\exp(\gamma_i) + \exp(\gamma_j)} \\ &= \frac{1}{1 + \exp\left(-(\gamma_i - \gamma_j)\right)} \\ &= \sigma(M(i, j)), \end{aligned} \tag{1}$$

where $\sigma(x) = 1/(1 + \exp(-x))$ is the sigmoid or logistic function.

The three core properties are: $M(i, j) > 0$ means i has more than 50% chance to win, and $M(i, i) = 0$ means it is an even matchup; $M(i, j) \to +\infty$ means $\Pr(i \succ j) \to 1$; the matchup matrix $\mathbf{M}$ satisfies negative symmetry, i.e.,

$$\mathbf{M} = -\mathbf{M}^\top. \tag{2}$$

Note that the negative symmetry property $M(i, j) = -M(j, i)$ (2) is put in place to ensures $p_{ij} + p_{ji} = 1$.

In the Bradley-Terry model, each player has a one-dimensional strength parameter and the probability of player i winning a match against j only depends on the relative advantage of player i over player j.

C. Blade-Chest model

A critical limitation of the Bradley-Terry model is that it assumes transitive relations among players; that is, if player $\gamma_i > \gamma_j$ (i.e., i has an advantage over player) and $\gamma_j > \gamma_k$, then $\gamma_i > \gamma_k$ holds. In other words, all players are constantly ordered. However, we can see such assumption does not hold in many applications. A simplest counter example is Rock-Paper-Scissors game where the rule is Paper $\succ$ Rock, Rock $\succ$ Scissor and Scissor $\succ$ Paper. Such relations have a property called intransitivity, which is defined as the following:

Definition III.1. *Matchup relations of n players contain (stochastic) intransitivity if there exist three players i, j and k such that*

- $\Pr(i \succ j) > 0.5$;
- $\Pr(j \succ k) > 0.5$;
- $\Pr(k \succ i) > 0.5$.

Chen and Joachims [16] proposed the Blade-Chest model that allows intransitive relations among players by introducing two extra D-dimensional vectors for each player i: $\mathbf{x}_i^{\text{blade}}$ and $\mathbf{x}_i^{\text{chest}}$. The matchup matrix of the Blade-Chest model[1] is given as

$$M_{ij} = \mathbf{x}_i^{\text{blade}\top} \mathbf{x}_j^{\text{chest}} - \mathbf{x}_j^{\text{blade}\top} \mathbf{x}_i^{\text{chest}} + \gamma_i - \gamma_j. \tag{3}$$

The Blade-Chest model is a multi-dimensional extension of the Bradley-Terry model and the multi-dimensional representation allows intransitive relations among players. Generally, the more dimensions the representation has, the more intransitivity the model allows.

A neural network framework of the Blade-Chest model has been proposed by Chen and Joachims [15]. The top layer of it is the blade-chest-inner model (3). The bottom layer uses a fully-connected feed-forward mapping linking the blade/chest vectors and feature vectors. In their two models called CONCAT and SPLIT, game feature vectors are contained along with the player features.

D. Blade-Chest-Sigma model

Duan et al. [17] proposed a generic formulation of the comparison model. They also assume a d-dimensional representation $\mathbf{x}_i \in \mathcal{R}^d$ for player i, then the matchup matrix is given by

$$M_{ij} = \mathbf{x}_i^\top \Sigma \mathbf{x}_j + \mathbf{x}_i^\top \Gamma \mathbf{x}_i - \mathbf{x}_j^\top \Gamma \mathbf{x}_j, \tag{4}$$

where $\Sigma, \Gamma \in \mathcal{R}^{d \times d}$ are the transitive matrices. $\mathbf{x}_i^\top \Sigma \mathbf{x}_j$ reflects the interaction between players, and $\mathbf{x}_i^\top \Gamma \mathbf{x}_i - \mathbf{x}_j^\top \Gamma \mathbf{x}_j$ reflects the intrinsic strength of each individual. We denote this model as "Blade-Chest-Sigma".

Specially, if taking the matrix Σ with

$$\Sigma = \begin{pmatrix} \mathbf{0} & \mathbf{I} \\ -\mathbf{I} & \mathbf{0} \end{pmatrix},$$

the generalized model (4) reduces to the Blade-Chest-Inner.

[1]More precisely, this model is called the Blade-Chest-Inner model, and they also propose another variant called the Blade-Chest-Dist model; however, there is no significant difference between them. And in practice, the Blade-Chest-Inner always achieves better prediction than Blade-Chest-Dist. Then we focus on the former in this paper.

IV. Intransitivity model

In this section, we generalize the Blade-Chest intransitivity model by a low-rank matrix, and propose our framework for the generalized model. We first detail the model and analysis some properties which can unify several existing models, and then we briefly introduce the neural network framework.

A. Generalized intransitivity model

We first give a generalized representation of the matchup matrix of the Blade-Chest model. Let us denote the representations of player i by

$$\mathbf{x}_i = \begin{pmatrix} \mathbf{x}_i^{\text{blade}} \\ \mathbf{x}_i^{\text{chest}} \end{pmatrix},$$

and define the representation matrix as

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}^{\text{blade}} \\ \mathbf{X}^{\text{chest}} \end{pmatrix},$$

where

$$\mathbf{X}^{\text{blade}} = (\mathbf{x}_1^{\text{blade}}, \ldots, \mathbf{x}_N^{\text{blade}}),$$
$$\mathbf{X}^{\text{chest}} = (\mathbf{x}_1^{\text{chest}}, \ldots, \mathbf{x}_N^{\text{chest}}).$$

We also denote the strength parameters in the original Bradley-Terry model by

$$\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_N).$$

Using the above notations, we can see the Blade-Chest-Inner model (3) can be represented as

$$\mathbf{M} = \mathbf{X}^\top \begin{pmatrix} \mathbf{0} & \mathbf{I} \\ -\mathbf{I} & \mathbf{0} \end{pmatrix} \mathbf{X} + \boldsymbol{\gamma}^\top \mathbf{1} - \mathbf{1}^\top \boldsymbol{\gamma} \tag{5}$$

$$= \mathbf{X}^{\text{blade}\top} \mathbf{X}^{\text{chest}} - \mathbf{X}^{\text{chest}\top} \mathbf{X}^{\text{blade}} + \boldsymbol{\gamma}^\top \mathbf{1} - \mathbf{1}^\top \boldsymbol{\gamma}. \tag{6}$$

Now we replace the matrix product $\mathbf{X}^{\text{blade}\top} \mathbf{X}^{\text{chest}}$ by a new matrix $\mathbf{Y}$ as

$$\mathbf{Y} = \mathbf{X}^{\text{blade}\top} \mathbf{X}^{\text{chest}},$$

which results in a general representation of the matchup matrix as

$$\mathbf{M} = \left(\boldsymbol{\gamma}^\top \mathbf{1} - \mathbf{1}^\top \boldsymbol{\gamma}\right) + \left(\mathbf{Y} - \mathbf{Y}^\top\right) \quad \text{s.t. } \operatorname{rank}(\mathbf{Y}) \leq D. \tag{7}$$

The rank D controls the intransitivity level that the model allows. By increasing the rank of $\mathbf{Y}$ or ultimately, by removing the rank constraint, we can represent an arbitrarily complex matchup matrix by $\mathbf{Y} - \mathbf{Y}^\top$, which is also implied by the existing research [16, Theorem 1].

B. Properties of the generalized intransitivity model

Firstly, it has been mentioned above that $\mathbf{Y} - \mathbf{Y}^\top$ can represent an arbitrarily complex matchup matrix by removing the rank constraint.

Secondly, it is important to notice that the model can still represent the intransitivity even when we reduce the rank of $\mathbf{Y}$ to 1. If we take a rank-1 matrix as $\mathbf{Y}$, that is,

$$\mathbf{Y} = (x_1^{\text{blade}}, x_2^{\text{blade}}, \ldots, x_N^{\text{blade}})^\top (x_1^{\text{chest}}, x_2^{\text{chest}}, \ldots x_N^{\text{chest}}),$$

the matchup matrix without the strength terms $\gamma^\top \mathbf{1} - \mathbf{1}^\top \gamma$ becomes

$$M_{ij} = x_i^{\text{blade}} x_j^{\text{chest}} - x_i^{\text{chest}} x_j^{\text{blade}}.$$

Assume that $i \succ j$ and $j \succ k$ (i.e., $M_{ij} > 0$ and $M_{jk} > 0$), then taking $x_i^{\text{chest}} > 0$, $x_j^{\text{chest}} < 0$, and $x_k^{\text{chest}} > 0$ shows $k \succ i$ (i.e., $M_{ik} < 0$), which gives the intransitive relations among three players in the rock-paper-scissors game.

Lastly, noting that $\mathbf{Y} = \mathbf{0}$ if $\text{rank}(\mathbf{Y}) = 0$, the model with the strength terms (7) is equivalent to the ordinary Bradley-Terry model when $\text{rank}(\mathbf{Y}) = 0$.

C. Model estimation

We now introduce our framework for the generalized intransitivity model. Initially, for a pairwise (i, j), we have two 0-1 vectors that encode the players' identities. The bottom is an embedding that link the player vector into an embedded vector. Assume to link the embedding vector and the final objective Y_{ij} and Y_{ji} by the linear mapping, which shows that

$$\begin{aligned} Y_{ij} &= A \begin{pmatrix} \mathbf{x}_i \\ \mathbf{x}_j \end{pmatrix} + b, \\ Y_{ji} &= A \begin{pmatrix} \mathbf{x}_j \\ \mathbf{x}_i \end{pmatrix} + b. \end{aligned} \tag{8}$$

For this linear transformation, the rank of matrix $\mathbf{Y}$ is consistence to the dimension in the framework. Besides, it is easy to show the negative symmetry of the matchup matrix as that

$$\begin{aligned} M_{ij} &= Y_{ij} - Y_{ji} = A \begin{pmatrix} \mathbf{x}_i \\ \mathbf{x}_j \end{pmatrix} - A \begin{pmatrix} \mathbf{x}_j \\ \mathbf{x}_i \end{pmatrix}, \\ M_{ji} &= Y_{ji} - Y_{ij} = A \begin{pmatrix} \mathbf{x}_j \\ \mathbf{x}_i \end{pmatrix} - A \begin{pmatrix} \mathbf{x}_i \\ \mathbf{x}_j \end{pmatrix}, \end{aligned}$$

and thus $M_{ij} = -M_{ji}$.

As a special case, if we define the mapping $A = \begin{pmatrix} B \\ C \end{pmatrix}$ and $b = 0$, the formulation (8) can be viewed as the linear transformation to blade and chest vectors in [15].

Alternatively and generally, we can use a fully-connected nonlinear neural network layer with the activation function f. The first hidden layer is given as an example that

$$\begin{aligned} z_{11} &= f_1(\mathbf{x}_i, \mathbf{x}_j), \\ z_{12} &= f_1(\mathbf{x}_j, \mathbf{x}_i). \end{aligned}$$

The top layer is the winning probability of the generalized model that

$$p_{ij} = \sigma(M(i, j)) = \sigma(Y_{ij} - Y_{ji}). \tag{9}$$

The entire pipeline is depicted in Figure 1.

In essence, this proposed framework is simpler than the framework of the Blade-Chest model in [15], because the "blade" and "chest" vectors are not explicitly updated in the learning process. Alternatively, the vectors are implicitly embedded in the model via the embedding modules and an explicit rank constraint.

V. Experiments

In this section, we evaluate the proposed model by using one synthetic dataset and several real-world datasets that range from food preference, recommender system and online games. We investigate the model by the accuracy of model in predictive setting.

A. Experiment settings

We first introduce the general setup of the experiments. The input of all experiments includes: a set of players $P = \{1, 2, ..., N\}$; a dataset $\mathcal{D}$ of all the match results between player i and j in P, for any $i, j \in P$, n_i be the number of times player i win j and n_j be the opposition. The output of the experiments is a estimated strength matrix $\mathbf{Y}$.

Assuming i is the winner, the objective function is

$$\text{argmax}_\Theta \ \Sigma_{(i,j)\in\mathcal{D}} \log \Pr(i \succ j | \Theta).$$

For a test partition $\mathcal{D}'$, the test accuracy is defined as

$$A(\mathcal{D}'|\mathbf{Y}) = \frac{1}{N'} \sum_{(i,j)\in\mathcal{D}} \mathbf{1}(i \succ j),$$

where N' is the total number of games in the testing set, and $\mathbf{1}(\cdot)$ is the indicator function. The model performs better when the accuracy value is higher.

For the baselines, we compared our proposed method with four competitive methods as following

1). Bradley-Terry model with the stochastic gradient method (BT model);
2). Blade-Chest-Inner with the stochastic gradient method (Blade-Chest-Inner);
3). Blade-Chest-Sigma with the stochastic gradient method (Blade-Chest-Sigma);
4). Blade-Chest-Inner with neural network framework (Neural BC).

We follow the same setting for BT model, Blade-Chest-Inner and Blade-Chest-Sigma with the SGD methods as those in [16], [17]. The objective functions are to optimize the maximum log-likelihood function with different regularization terms. It includes three parameters: the parameter λ of the regularization term, the learning rate r and the embedding dimension d. We do grid search over powers of 10 from 1e-5 to 1e$+2$ for λ and r. We take $d = \{2, 5, 10, 50, 100\}$. For the Neural BC and our proposed method, we take the batch size from 16 to 1024 and the middle dimension from 8 to 512.

B. Experiments with synthetic datasets

In this experiment, the simple synthetic datasets are firstly used for performance evaluations. Next, we also test the proposed algorithm with several real-world data to demonstrate the expressiveness.

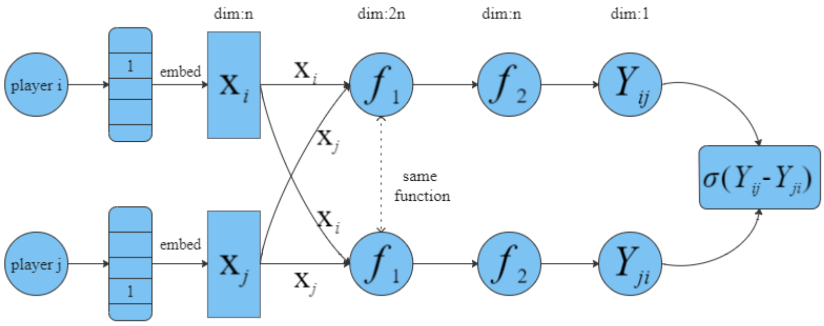

Figure 1. An illustration of the proposed generalized intransitivity framework

1) Synthetic datasets: We randomly generate the datasets which have the full intransitivity with different ranks and sizes. Let $R = \{1, 3, 5, 7\}$ be a set of truth ranks which means the dimensions of the blade and chest vectors, and $S = \{500, 1000, 2000\}$ be a set of training data size. Given $N = 100$ players, for every rank element $r \in R$, we first randomly generate blade and chest parameters $\mathbf{X}^*_{\text{blade}} \in \mathcal{R}^{r\times 100}$ and $\mathbf{X}^*_{\text{chest}} \in \mathcal{R}^{r\times 100}$. Then the relation matrices $\mathbf{Y}^*$ and $\mathbf{M}^*$ are calculated by

$$\mathbf{Y}^* = \mathbf{X}^*_{\text{blade}}{}^\top \mathbf{X}^*_{\text{chest}}, \quad \mathbf{M}^* = \mathbf{Y}^* - \mathbf{Y}^{*\top}.$$

Then for all sizes $s \in S$, we randomly generate a training data T with size s, a validation data V with size 2000 and a evaluation data E with size 2000 using the true relation matrix $\mathbf{M}$.

For each rank, there are three different training sets, and we generate 10 trials for each rank.

2) Experiment results: We train our model with all the baselines for these four rank groups and the accuracy results are averaged over the 10 trials which are shown in Figure 2.

Figure 2 shows the curves of the accuracy values with respect to the sizes of training datasets. For all the competitive methods, the accuracy value turns to be higher around 0.9 along with the bigger dataset, which also shows the better performance. The original intransitivity models Blade-Chest-Inner and Blade-Chest-Sigma perform nearly the same and better than the basic BT model, while our proposed new model outperforms them and is consistence to the neural Blade-Chest model. This indicates that our generalized intransitivity model is simple but has the same expressiveness as the multi-dimensional Blade-Chest model.

C. Experiments with real datasets

We now move on to the real-world applications to investigate the feasibility and effectiveness of our proposed method.

1) Real-world datasets: The datasets are generally used in pairwise comparison problems which includes food preference datasets: SushiA and SushiB [8], online recommender system for pairwise preference datasets: MovieLens100K [7], decision making datasets: Elections in terms of several sizes [10], and online games: SF4, Dota [16] and Pokemon.

The SushiA and Election A48 are not intransitivity, MovieLens100K and other Elections have lower ratios of intransitivity, while SushiB and online games are with a higher ratios. We list all the related intransitive relationships in Table I. Intrans. means the existence of the intransitivity, No.IntPlayer indicates the number of players those involved in rock-paper-scissors relationship, and the Int.Ration is the percentage of intransitive loops in the whole games.

We use 50% matches for training, 20% matches for validation and 30% matches for testing in each dataset by randomly separating. We do this random splitting for 3 times and report the mean and standard deviation.

2) Experiment results: In this subsection, the performances of all the comparison methods for real-world datasets are explored. Table II shows the averaged test accuracies.

Firstly, focusing on the comparison with baselines BT model, Blade-Chest-Inner and Blade-Chest-Sigma with SGD methods, the neural BC and our proposed model have relatively larger improvement on the test accuracy. It could tell that the neural network framework make a bigger contribution. Then we compare the new generalized model with the neural BC. Our new method outperforms the baseline on monst of the datasets, however it just leads to typically small improvements.

VI. Conclusions

In this paper, we presented a generalized intransitive framework for modeling pairwise comparison. We proposed a low-rank matrix formulation of pairwise matchups instead of conventional multi-dimensional vectors that were used to characterize each player. The proposed framework is efficient to address the intransitive relationships that are abnormal for Bradley-Terry model but widely exists in datasets. We also showed an unified perspective of our model and its linkage to the Bradley-Terry model, the Blade-Chest model and the Neural Blade-Chest model. The effectiveness of the proposed model is evidenced in the experiments on a synthetic dataset and various real-world datasets. The model outperforms most of the baseline methods and enhanced the accuracy in predictive setting. It is distinctly simpler than the neural Blade-Chest framework.

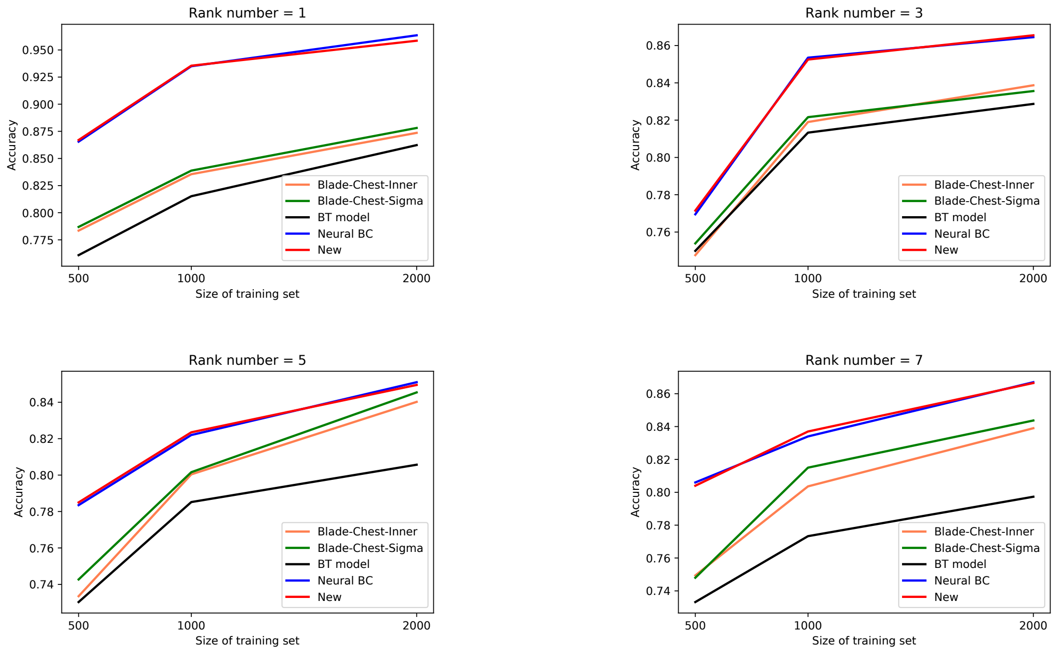

Figure 2. Test accuracy with the synthetic datasets

Table I
SUMMARY OF THE DATASETS

Dataset	Players	Records	Intrans.	No.IntPlayer	Int.Ratio
SushiA	10	100000	no	0	0
SushiB	100	25000	yes	92	26.87%
MovieLens100K	1682	139982	yes	1130	0.19%
Election A5	16	44298	yes	6	0.44%
Election A9	12	95888	yes	5	1.82%
Election A17	13	21037	yes	8	8.18%
Election A48	10	25848	no	0	0
Election A81	11	44298	yes	5	2.50%
SF4-5000	35	5000	yes	34	23.86%
Dota	757	10442	yes	550	97.58%
Pokemon	800	50000	yes	784	78.58%

Table II
TEST ACCURACY

Dataset	Bradley-Terry	Blade-Chest-Inner	Blade-Chest-Sigma	Neural BC	New model
SushiA	0.6525 $\pm$ 0.0011	0.6546 $\pm$ 0.0006	0.6560 $\pm$ 0.0004	0.6630 $\pm$ 0.0004	**0.6632 $\pm$ 0.0003**
SushiB	0.6257 $\pm$ 0.0025	0.6235 $\pm$ 0.0150	0.6414 $\pm$ 0.0019	0.6561 $\pm$ 0.0017	**0.6563 $\pm$ 0.0011**
MovieLens100K	0.6785 $\pm$ 0.0005	0.6792 $\pm$ 0.0004	0.6789 $\pm$ 0.0003	0.6950 $\pm$ 0.0019	**0.6973 $\pm$ 0.0002**
Election A5	0.6478 $\pm$ 0.0017	0.6489 $\pm$ 0.0011	0.6494 $\pm$ 0.0018	0.6550 $\pm$ 0.0030	**0.6560 $\pm$ 0.0018**
Election A9	0.6028 $\pm$ 0.0003	0.6096 $\pm$ 0.0007	0.6047 $\pm$ 0.0008	0.6174 $\pm$ 0.0003	**0.6175 $\pm$ 0.0003**
Election A17	0.5189 $\pm$ 0.0001	0.5305 $\pm$ 0.0010	0.5296 $\pm$ 0.0013	0.5582 $\pm$ 0.0003	**0.5598 $\pm$ 0.0002**
Election A48	0.5993 $\pm$ 0.0001	0.6001 $\pm$ 0.0001	0.5996 $\pm$ 0.0001	0.6060 $\pm$ 0.0001	**0.6056 $\pm$ 0.0001**
Election A81	0.6013 $\pm$ 0.0001	0.6018 $\pm$ 0.0001	0.6011 $\pm$ 0.0002	0.6194 $\pm$ 0.0001	**0.6194 $\pm$ 0.0001**
SF4-5000	0.5079 $\pm$ 0.0078	0.5181 $\pm$ 0.0171	0.5358 $\pm$ 0.0049	0.5514 $\pm$ 0.0008	**0.5496 $\pm$ 0.0021**
DotA	0.6334 $\pm$ 0.0077	0.6432 $\pm$ 0.0034	0.6420 $\pm$ 0.0051	0.6468 $\pm$ 0.0031	**0.6485 $\pm$ 0.0025**
Pokemon	0.8157 $\pm$ 0.0094	0.8495 $\pm$ 0.0016	0.8187 $\pm$ 0.0168	0.8943 $\pm$ 0.0040	**0.8949 $\pm$ 0.0021**

References

[1] T. Graepel, T. Minka, and R. T. Herbrich, "A Bayesian skill rating system," Advances in Neural Information Processing Systems, vol. 19, pp. 569–576, 2007.

[2] J. E. Menke and T. R. Martinez, "A Bradley–Terry artificial neural network model for individual ratings in group competitions," Neural computing and Applications, vol. 17, no. 2, pp. 175–186, 2008.

[3] M. Cattelan, C. Varin, and D. Firth, "Dynamic Bradley–Terry modelling of sports tournaments," Journal of the Royal Statistical Society: Series C (Applied Statistics), vol. 62, no. 1, pp. 135–150, 2013.

[4] M. Regenwetter, J. Dana, and C. P. Davis-Stober, "Transitivity of preferences." Psychological review, vol. 118, no. 1, p. 42, 2011.

[5] D. F. Gleich and L.-h. Lim, "Rank aggregation via nuclear norm minimization," in Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, 2011, pp. 60–68.

[6] M. Jamali and M. Ester, "A transitivity aware matrix factorization model for recommendation in social networks," in Twenty-Second International Joint Conference on Artificial Intelligence, 2011.

[7] J. L. Herlocker, J. A. Konstan, A. Borchers, and J. Riedl, "An algorithmic framework for performing collaborative filtering," in ACM SIGIR Forum, vol. 51, no. 2. ACM New York, NY, USA, 2017, pp. 227–234.

[8] T. Kamishima, "Nantonac collaborative filtering: recommendation based on order responses," in Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, 2003, pp. 583–588.

[9] L. L. Thurstone, "A law of comparative judgment," Psychological review, vol. 34, no. 4, p. 273, 1927.

[10] N. Tideman, Collective decisions and voting: the potential for public choice. Routledge, 2017.

[11] R. A. Bradley and M. E. Terry, "Rank analysis of incomplete block designs: I. the method of paired comparisons," Biometrika, vol. 39, no. 3/4, pp. 324–345, 1952.

[12] R. D. Luce, Individual choice behavior: A theoretical analysis. Courier Corporation, 2012.

[13] M. Cattelan, "Models for paired comparison data: A review with emphasis on dependent data," Statistical Science, pp. 412–433, 2012.

[14] D. Causeur and F. Husson, "A 2-dimensional extension of the Bradley–Terry model for paired comparisons," Journal of statistical planning and inference, vol. 135, no. 2, pp. 245–259, 2005.

[15] S. Chen and T. Joachims, "Predicting matchups and preferences in context," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 775–784.

[16] ——, "Modeling intransitivity in matchup and comparison data," in Proceedings of the ninth acm international conference on web search and data mining. ACM, 2016, pp. 227–236.

[17] J. Duan, J. Li, Y. Baba, and H. Kashima, "A generalized model for multidimensional intransitivity," in Pacific-Asia Conference on Knowledge Discovery and Data Mining. Springer, 2017, pp. 840–852.

[18] R. R. Davidson, "On extending the Bradley-Terry model to accommodate ties in paired comparison experiments," Journal of the American Statistical Association, vol. 65, no. 329, pp. 317–328, 1970.

[19] M. E. Glickman, "A comprehensive guide to chess ratings," American Chess Journal, vol. 3, no. 1, pp. 59–102, 1995.

[20] P. Linares, "Are inconsistent decisions better? an experiment with pairwise comparisons," European Journal of Operational Research, vol. 193, no. 2, pp. 492–498, 2009.

[21] Z. Cao, T. Qin, T.-Y. Liu, M.-F. Tsai, and H. Li, "Learning to rank: from pairwise approach to listwise approach," in Proceedings of the 24th international conference on Machine learning, 2007, pp. 129–136.

[22] T.-K. Huang, C.-J. Lin, and R. C. Weng, "Ranking individuals by group comparisons," Journal of Machine Learning Research, vol. 9, no. Oct, pp. 2187–2216, 2008.

[23] J. P. Keener, "The Perron–Frobenius theorem and the ranking of football teams," SIAM review, vol. 35, no. 1, pp. 80–93, 1993.

[24] R. Herbrich, T. Minka, and T. Graepel, "Trueskill^{TM}: A bayesian skill rating system," in Advances in Neural Information Processing Systems 19. MIT Press, 2007, pp. 569–576.

[25] P. Dangauthier, R. Herbrich, T. Minka, and T. Graepel, "TrueSkill through time: Revisiting the history of chess," in Advances in neural information processing systems, 2008, pp. 337–344.

[26] L. Zhang, J. Wu, Z.-C. Wang, and C.-J. Wang, "A factor-based model for context-sensitive skill rating systems," in 2010 22nd IEEE International Conference on Tools with Artificial Intelligence, vol. 2. IEEE, 2010, pp. 249–255.

[27] R. P. Adams, G. E. Dahl, and I. Murray, "Incorporating side information in probabilistic matrix factorization with Gaussian processes," arXiv preprint arXiv:1003.4944, 2010.

[28] D. R. Hunter et al., "Mm algorithms for generalized Bradley-Terry models," The annals of statistics, vol. 32, no. 1, pp. 384–406, 2004.

[29] P. Gopalan, J. M. Hofman, and D. M. Blei, "Scalable recommendation with poisson factorization," arXiv preprint arXiv:1311.1704, 2013.

[30] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in Advances in neural information processing systems, 2013, pp. 3111–3119.

[31] G. Adomavicius, R. Sankaranarayanan, S. Sen, and A. Tuzhilin, "Incorporating contextual information in recommender systems using a multidimensional approach," ACM Transactions on Information Systems (TOIS), vol. 23, no. 1, pp. 103–145, 2005.

[32] L. Baltrunas and F. Ricci, "Context-based splitting of item ratings in collaborative filtering," in Proceedings of the third ACM conference on Recommender systems, 2009, pp. 245–248.

[33] S. Rendle, Z. Gantner, C. Freudenthaler, and L. Schmidt-Thieme, "Fast context-aware recommendations with factorization machines," in Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval, 2011, pp. 635–644.

[34] J. Huber, J. W. Payne, and C. Puto, "Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis," Journal of consumer research, vol. 9, no. 1, pp. 90–98, 1982.

[35] I. Simonson and A. Tversky, "Choice in context: Tradeoff contrast and extremeness aversion," Journal of marketing research, vol. 29, no. 3, pp. 281–295, 1992.

Acknowledgment

This work was partially supported by JSPS KAKENHI Grant Number 15H01704 and 20H04244.