Intransitivities in videogames and matchmaking

CE888 Assignment 1

Raul Axel Suarez Martinez
MSc Student at University of Essex.
Colchester, Essex, UK.
rs16200@essex.ac.uk

Abstract**—** In this paper it is discussed an investigation about intransitivities and how they affect the match ranking. This kind of systems have their base around 30 years ago and have been a success in disciplines like chess, but for other types of contests were several factors are taken into account, there could be the possibility of presenting an intransitivity. This project will try to reinforce the idea that if state-of-the-art ranking methods could consider more than one dimension, the performance of such classifiers will improve.

Keywords—intransitivity; ranking; matchmaking.

I. Introduction

Pairwise comparison or matchmaking (used interchangeably later in the text) is the process of comparing to agents in order to select one based on the quantitative value of a feature. In the area of multiagent systems, it could be used to select the agent with most probability of winning a contest. This is done by modeling the skill to get a value of an agent and compare it with others through a selector, but in more complex scenarios the difficulty of intransitivities arise. An intransitivity is present in a graph where A → B and B → C but C → A, a clear example of this being the game Rock, Paper, Scissors. Because most of the pairwise comparison methods doesn't take into account this intransitivities they are prone to predict a low probable outcome for some matches when an unconsidered feature of the lowly ranked agent make it particularly stronger against its opponent. This is a case with often appearance in the field of multiplayer online videogames, were game designers plan this intrasitivities as a way of balancing the game.

II. Background

The topic of pairwise comparison has been broadly discussed previously, one of the models that found a direct impact in application and is currently the most used is the Elo ranking [1]. This ranking is the base for numerous ranking models and follows the idea that any given player i that show performance $p_i \sim N(p_i;\, s_i, \beta^2)$, and the probability that any given player, lets say player p_1 to uotperform player p_2 is the probability that the performance of $p_1 > p_2$:

$$P(p_1 > \ p_2 \mid s_1, s_2\ ) = \ \Phi\ (\frac{s_1 - s_2}{\sqrt{2}\beta})$$

In his work, Elo states that the distribution for the probability for an agent to show its strenght in a match followed a normal distribution, although some other methods use a logistic distribution. Given the positive results of this ranking in chess, the searching for new models which can achieve similar or better performance have been apiering through the years. In [2] they present a bayesian approach to generalize the Elo ranking through factor graphs and using the logisticdistribution, achieveing improved performance over the baseline and testing in a competitive videogame environment. This approach is commercially used for the videogame industry.

These methods however, rely solely on a single scalar to describe the ability of a competitor for a given discipline or game, and there are given circumstances where this does not describe accurately the strengths or weakness that can arise on a particular match.

In [3] they present a learning model which can potentially learn intransivity preference relations in a multidimentional space for pairwise comparison problems. That is one of the few methods which take into account more than single scalar, and they are base on the Bradley-Terry model. Unfortunately, they do not focus on identifying intransivities in the matches but assumed that this intransitivities exists to validate their research of modeling skill in multiple dimensions.

A work conducted in coevolution algorithms that does measure intransitivities [4] provide the metrics and their implementations to measure intransitivity. In this work, they propose 2 ways of doing the measurement: with a transitivity index and with a KL Divergence. The first one consists in a simple index of the number of 3-dags in a tournament divided by the number of posible 3-dags, where a 3-dag is a transitive graph between 3 agents:

$$\tau : \frac{T_3}{\binom{n}{3}}$$

Althouhgt the simplicity of the method, is straightforward to analyse this measurement: the higher it is to 1, there are less intransitivities. As the authors specified, the problem with this index is that it highly reactive and the minor change in the tournament results could affect significantly the number. This is why the second approach is also used, because an stadistical measure could lead to more stable results.

III. Methodology

Through several experiments I'll try to analyze the number of intransitivities that can happen in a given set of matchups, if they are representative to the sample and how do they affect the ranking system. Several questions can arise during the experimentation part which could lead the analysis to different paths:

- Do the current ranking system used in the dataset truly reflect the probability of an agent to win over another agent?
- How big is the ranking difference for those intransitivity cases?
- Does the matchmaking system has a strong type of intransitivity?

Trying to find an answer for these questions would take a bigger amount of time and resources than those expected for the scope of this project, but they could be the starting point for a future further analysis. For the matter of this project, I will focus on finding how does the intransitivities (if any) in the dataset affect the matchmaking by identifying in which sections of the ranking does the majority of the intransitivities occur.

The dataset I will use for the experiments is the StarCraft Alilugac database [5] updated to the 22 of February 2017, which holds records of 400K games and 200K matches of competitive matchups as well as event and player information. StarCraft is an online real-time strategy game created by Blizzard Entertainment [6] which can be played individually or multiplayer format. Because of simplification purposes, I will examine the data related only to single-player 1v1 matches. A visualization of the matches table structure extracted from the aligulac site can be found below:

Name	**Definition**	**Description**
id	integer not null	primary key
period_id	integer not null	foreign key to period (the period this match was played in)
date	date not null	when the match was played (often approximate)
pla_id	integer not null	foreign key to player (player A)
plb_id	integer not null	foreign key to player (player B)
sca	smallint not null	score for player A
scb	smallint not null	score for player B
…		
rta_id	integer	foreign key to rating (rating of player A at the time of the match)
rtb_id	integer	foreign key to rating (rating of player B at the time of the match)

Table 1 Structure of the first columns of the matches table

As the table shows, the information about the players, the scores and the ranking is provided. The ranking is an id due to the rankings are not the officials by Blizzard but ones generated by the creators of the database and they change through time so a table of only rankings is generated, the method of such rankings is the Glickman method [7]. Because of the size of the dataset, a process of sampling will be performing to use only a portion of the whole data and have faster tests.

The technology selected for the project is python 2 with the help of pandas for data manipulation.

IV. Experiments

The pipeline to follow on this project is the next one:

1. Extraction and sampling: load a random sample of the values from the database into memory.
2. Pre-processing / preparation: manipulating the data frames to merge scores with matches and dropping extra columns.
3. Identify intransitivities: form 3 nodes graphs to find and save the number of intransitivities.
4. Create a Bradley-Terry model predicter based on the rankings and train it, try to predict the matches
5. Compute metrics: calculate the Transitivity Index and the KL Divergence.
6. Analyze results.

The most important steps of the process will be 3, 4 and 5. In step 3 the intransitivities are recorded to later get the Transitivity Index, which is necessary to know the proportion of intransitivities in the matches, where 4 is necessary to get the KL Divergence. For the model, I will use both a normal and a probabilistic distribution, to find out which one is the best for this kind of data.

The experiments, given the results of the previous pipeline, can vary depending on the outcome of the fifth step. There is the possibility of mapping the rankings to the intransitivities identified to know in which section lies the higher concentration of intransitivities.

In the case of having a low performance on the predictor, an Elo ranking values would be calculated from scratch for every player on the sample, and then apply the rest of the pipeline to those results. This could be done to identify if the ranking given from the website is the problem or the data indeed is full of unexpected results.

Lastly, the whole process will be done from 5 to 10 times (depending on time available) with different samplings from the dataset to validate that the results were not sampling correlated.

V. Discussion

Because this is a project of analysis and not heavily focused on implementation, the use of metrics for directly comparing results is not possible on every output of the pipeline. The main conclusions of the work would be drawn from comparing the observations of the Transitivity index and the KL Divergence.

To measure and compare the performance of the Bradley-Terry predicter, a random approach predictor will be used as baseline, in the case of underperforming, the data of the classification based on the Elo ranking also will be attached to the comparison. And the heuristic to determine high intransitivity for the Transitivity Index will be the arbitrary seleced value of $\tau \leq 0.5$. Is important to also mentioned that after having the analysis of the performance of traditional ranking methods by their proportion of intransitivities, a good suggestion for future topic research would be a comparison of ranking systems and their performance in this dataset, with the approach of [].

VI. Conclusion

In conclusion, this assignment will try to identify intransitivities and how they affect the ranking in matchmaking particularly in the game of Starcraft. Based on the system proposed by Elo and the Bradley-Terry model for pairwise comparison, a predictor will be build in order to perform an statistical driven method to measures intransitivities, the KL Divergence. At the same time, an analytical method called Transitivity Index will serve as an indicator of the ratio between transitive and intransitive relations. The conclusions drawn by this experiments could potentially incentive new methods of ranking classification with more than one dimension, to consider strong intransitive relations which make some agents highly favorable against anothers.

VII. Plan

In order to organize the implementation of the project, a table with the estimated efforts will be provided:

Step	**Start Date**	**Estimated time**
Extraction and sampling	3/3/2017	1/2 day
Pre-processing / preparation	10/3/2017	2 days
Identify intransitivities	26/3/2017	3 days
Create a Bradley-Terry model predicter	29/3/2017	5 days
Compute metrics	4/4/2017	2 days
Analyze the results	7/4/2017	3 days
Writing assignment report	12/4/2017	3 days

Table 2 Estimated schedule for the assignment development

VIII. References

[1] A. Elo, The Rating of Chess Players, Past and Present, Ishi Press, 2008.

[2] R. Herbrich, T. Minka and T. Graepel, "True Skill: A Bayesian Skill Rating System," *Proceedings of the 19th International Conference on Neural Information Processing Systems*, pp. 569-576, 2006.

[3] S. Chen and J. Thorsten, "Modeling Intransitivity in Matchup and Comparison Data," *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining* , pp. 227-236, 2016.

[4] S. Samothrakis, S. Lucas, T. P. Runarsson and D. Robles, "Coevolving Game-Playing Agents: Measuring Performance and Intransitivities," *Spyridon Samothraki*, vol. 17, pp. 213-226, 2013.

[5] "Aligulac," [Online]. Available: http://aligulac.com/about/db/. [Accessed 22 February 2017].

[6] "Starcraft 2 Battle . NET," Blizzard Entertainment, [Online]. Available: http://us.battle.net/sc2/en/game/. [Accessed 22 February 2017].

[7] M. E. Glickman, "Parameter estimation in large dynamic paired comparison experiments," *Applied Stadistics*, pp. 377-394, 1999.