

CS4211 Formal Methods for Software Engineering

Default Project: Applying Probabilistic Model Checking in Sports Analytics

Introduction:

Traditionally, Probabilistic Model Checking (PMC) was used to analyze the correctness and performance of computer systems and protocols. However, due to the expressiveness and generality of PMC's capabilities, it can be applied in other areas as well. For example, during the lecture, we demonstrated an interesting case where PMC is applied to tennis analytics to reason about the relationship between a player's strategy and his winning probabilities. This example models a tennis singles tiebreaker game. Using this model, we are able to predict who will win the match and suggest player strategies to improve their chances of winning. Similarly, we have developed a soccer model to predict the winner in the English Premier League. We invite you to **apply the PMC** technique to **extend the tennis or soccer model**, or to create a model for a basketball game (NBA). Through this project, we hope you gain experience in modeling a realistic system using CSP# (PAT).

To help you, we would like to offer the following tips:

- Use **proper abstraction** to model the states, the state transitions and the player's choices. We have provided tennis and soccer models as examples.
- **Part of the challenge is to estimate the probabilities** in your model. We have listed data sources and scripts for this purpose.
- Your model will **be used to predict the winner of the match**. We will provide betting simulation scripts, so that you can compare your model's performance with our example model. You can treat the betting simulation as a sanity check, that is models with reasonable prediction performance should have similar performance compared to our example model.
- **Disclaimer**, please note that the betting simulation is intended exclusively for evaluating your model's predictive capabilities. This is based on the recognition that bookmakers are widely recognized for their accurate match result predictions. It is important to clarify that we do not endorse any illegal betting activities and cannot be held responsible for any legal repercussions or financial losses resulting from your engagement in any form of betting.
- One overall project (team) will be awarded a certificate of "Best Project of CS4211".

Data sources and model examples:

Tennis

Original datasource: <https://www.tennisabstract.com/charting/meta.html>

Data in CSV format:

https://drive.google.com/file/d/1pHo8PfkGxdgjsHljUf9poEF_vxhdzTA0/view

Tennis PAT model: Refer to tennis model example explained during the lecture, additionally, you can refer to these 6-region example models, where each side of the court is divided into 3 regions, that is left/middle/right regions. https://drive.google.com/drive/folders/1iy_X-jw4BwfdMsnlrW82sNvXzKjEbJhX

Elo ranking calculation: https://github.com/jgollub1/tennis_match_prediction/tree/master

Potential ideas to extend the existing model:

1. Model multiple shots strategy by including previous 2 shots
2. Modeling more types of shots

Tutor for tennis project: Dr. JIANG KAN <jiangkan@nus.edu.sg>

Soccer

Dataset + PAT Model Example:

https://drive.google.com/drive/folders/1Bm_nnJALqkOZdt1MTwQJZORmZAt8xmZs?usp=sharing

A model similar to tennis was done for soccer (the example code for it is heavily commented). The probabilities for the actions each player can take were taken from the FIFA video game. Some abstractions were used to simplify the model:

1. Only evaluate one team attacking and defending at any time (2 models for every match)
2. Fix player positions according to abstracted formations
3. Ball only moves forward
4. Stop trace once intercepted

Potential ideas to extend the existing model:

1. Include more actions that are in the FIFA ratings dataset
2. Change the model structure to something entirely different

Tutor for soccer project: Rajdeep Singh Hundal <rajdeep@u.nus.edu>

Basketball

The task of basketball modeling is both novel and intricate. Presently, we do not have any example model. Nonetheless, our tutors are committed to aiding you in the development of your own model. We strongly urge you to utilize the tutorial sessions to consult the tutors and address any inquiries you may have.

Datasource 1: https://github.com/swar/nba_api

Datasource 2: Wowy stats <https://github.com/shanefenske/nba-shots-wowy/blob/8bc2bd4245e304128742d17799219e7e45333adc/README.md>

Datasource 3: Pbpstats api <https://api.pbpstats.com/docs#/>

Tutor for basketball project: Liu Zhaoyu <e0253678@u.nus.edu>

Timeline and Assessment:

- Forming a team (in week 3 or week 4, group size can be 1, 2, 3, or 4)
- Mid term presentation: 10% (5 Oct 2023)
- Final project presentation: 20% (9 Nov 2023)
- Final project report: 30% (27 Nov 2023)