

Sports Match Outcome Prediction with Spatio-Temporal Graph Representation Learning

Abdolreza Mirzaei
Simon Fraser University
abdolreza_mirzaei@sfu.ca

Oliver Schulte
Simon Fraser University
oschulte@cs.sfu.ca

Mohammad Bahrami
Isfahan University of Tech.
mohammadbahrami@ec.iut.ac.ir

Masoud Mousavi
Isfahan University of Tech.
mmousavi@ec.iut.ac.ir

Abstract

Predicting the outcome of a sports match is a widely studied challenging problem of interest to many stakeholders (media, teams, fans, bookmakers). This paper develops a novel approach to match outcome prediction that leverages graph representation learning to model team-team and team-player interactions. Both teams and players correspond to nodes in a Spatio-Temporal graph. Node embeddings capture how team/player characteristics jointly influence match outcomes. Empirical results on a dataset of nine different football leagues demonstrate the superior performance of our graph representation approach.

Problem Definition Basically, sports match prediction refers to predicting future outcomes based on past results and features. So in general any observation will include the information of the two teams playing at a certain time and the result of that match. There are two different types of match predictions: *pre-match* and *live (ongoing)* prediction. In this paper, predictions are made pre-game. The main purpose of this paper is to predict the outcome of sports matches to {win, lose, draw} by only using dynamic graph (specifically Spatio-Temporal graph) representation learning and without using any hand-crafted or sport-specific features. Another key contribution of this paper is the use of lineup information.

Related Work In recent years, there has been much research on the use of machine learning in the prediction of match outcomes. These studies have mainly been based on feature engineering and have considered only a small number of leagues. Various machine learning methods including Neural Networks have been used in this field [5][9][8][2][1][4]. Recently, deep neural networks have been used in a few studies [3][6][7], and to the best of our knowledge, graph neural networks have been used in only one case [7]. The majority of past research has focused on learning about teams, not players. In other words, players' strength and lineup information, has not been considered in most previous works, despite their significant impact on match outcome prediction.

Methodology In competitive team sports, the outcome can depend on complex interactions between opposing teams and also interactions among players. Given that at a given time t , players, teams and their relations can be expressed by a spatial graph structure and the relationships between teams/players at different times can be expressed by a temporal structure, we propose a novel framework to predict match outcome leveraging *Spatio-Temporal graph representation learning*.

In the proposed Spatio-Temporal graph, each player and each team at a given time is assigned a node. An edge between teams represents a match outcome (e.g., from the winning team to the losing team), and an edge between a team and a player represents a player appearing for a team. In more detail, at each time step, players who have played for a team are connected to it via a *PlayedFor* link. At each time step, based on the result of every match, the two competing teams' nodes are connected with links of type *won* or *lost*, or *tied*. Every team or player node at time t is connected to its corresponding node at time $t - 1$ with the link of type *IsAfter*. These *IsAfter* links work as regularizers to enforce similarity between corresponding

nodes at different times and also encode the players and teams evolutions. It should be noted that for every relation, its inverse relation is also used with the with opposite directionality.

In the next step, by applying the graph representation learning methods to the graph of the teams/players and their relations, a representation is obtained for each node. Finally, the representations of the teams are used to predict the outcome of an upcoming match. An overall architecture of the introduced framework is depicted in figure 1. This approach provides several capabilities simultaneously:

1-Handling player learning can be achieved easily through this model, 2-There is no need for feature engineering and the learner can predict the match outcomes with raw input features (if any) and representation learning, 3-The evolution of players and teams over time becomes part of the Spatio-Temporal graph structure and does not require a separate solution, 4-The relational nature of the input data is well suited to this model. In addition, the process for adding any background information, if any, as a fact to the constructed graph is quite straightforward.

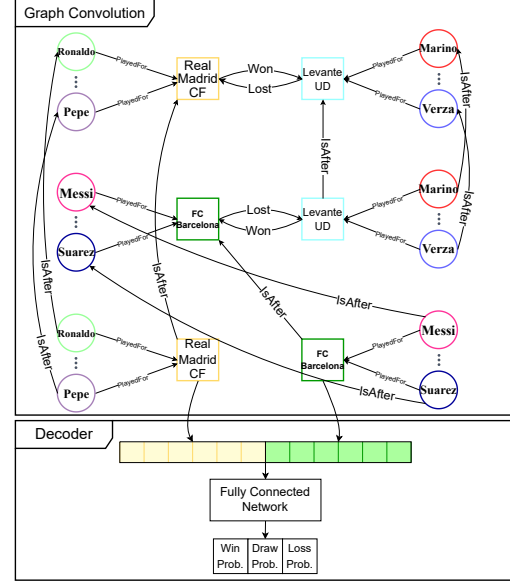


Figure 1: An illustration of the proposed framework for sport result prediction.

Results To evaluate the proposed method, we used soccer data from 9 leagues between 2008 and 2016 available at: <https://www.kaggle.com/hugomathien/soccer> The results of conducted experiments demonstrate that our framework outperforms other methods in almost all leagues in terms of accuracy and RPS metrics. Moreover, the visualization of the learnt representation of players displayed in figure 2, illustrates that the proposed model has been able to bring players with similar abilities closer together in the latent space. This result is interesting in that our model has no information about the number of goals, passes, dribbles or any other box score information about the players. There is only one input: the results of the matches in which they participated. To do an ablation study, we removed player nodes and player participation information from our method. Using our proposed method without this information, we found that the overall accuracy decreased from 50.36% to 49.63%, indicating that although the proposed framework still has a good ability to predict the outcome of the matches, participation information is still required.

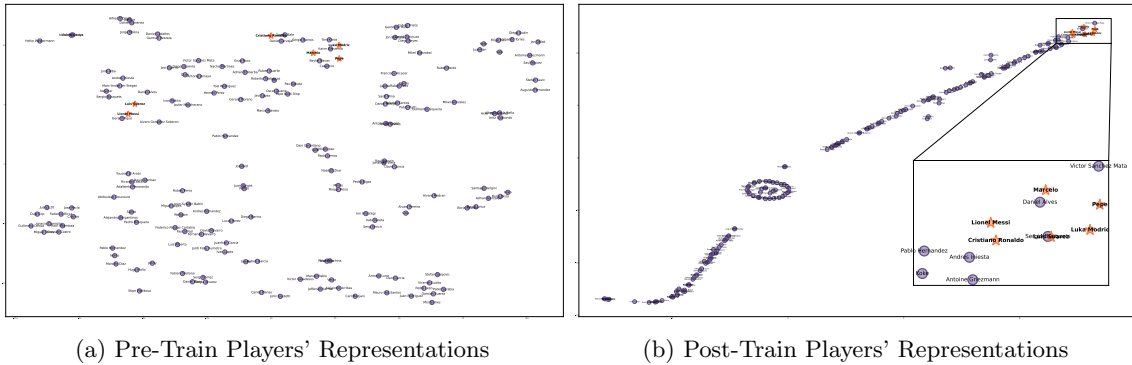


Figure 2: 2D representation of the Spain LIGA BBVA 2015-2016's last match week players obtained via the Pre-Train (a) and the Post-Train (b) models.

References

- [1] Daniel Berrar, Philippe Lopes, and Werner Dubitzky. Incorporating domain knowledge in machine learning for soccer outcome prediction. *Machine learning*, 108(1):97–126, 2019.
- [2] Anthony C Constantinou. Dolores: a model that predicts football match outcomes from all over the world. *Machine Learning*, 108(1):49–75, 2019.
- [3] Norbert Danisik, Peter Lacko, and Michal Farkas. Football match prediction using players attributes. In *2018 World Symposium on Digital Intelligence for Systems and Machines (DISA)*, pages 201–206. IEEE, 2018.
- [4] Ondřej Hubáček, Gustav Šourek, and Filip Železný. Learning to predict soccer results from relational data with gradient boosted trees. *Machine Learning*, 108(1):29–47, 2019.
- [5] Josip Hucaljuk and Alen Rakipović. Predicting football scores using machine learning techniques. In *2011 Proceedings of the 34th International Convention MIPRO*, pages 1623–1627. IEEE, 2011.
- [6] Robert Nyquist and Daniel Pettersson. Football match prediction using deep learning. Master’s thesis, Chalmers University of Technology, Gothenburg, Sweden, 2017.
- [7] Aleksandra Pereverzeva. Predicting sports matches with neural models. Master’s thesis, Czech Technical University in Prague, Prague, 2021.
- [8] Darwin Prasetio et al. Predicting football match results with logistic regression. In *2016 International Conference On Advanced Informatics: Concepts, Theory And Application (ICAICTA)*, pages 1–5. IEEE, 2016.
- [9] Niek Tax and Yme Joustra. Predicting the dutch football competition using public data: A machine learning approach. *Transactions on knowledge and data engineering*, 10(10):1–13, 2015.