TRANSFORMER-BASED MODEL TO PREDICT THE OUTCOMES OF SOCCER MATCHES FROM LEAGUES WORLDWIDE.

A PREPRINT

Tarak Kharrat
R&D Department
Real Analytics

Meher Kharbachi
R&D Department
Real Analytics

3/28/23

ABSTRACT

In this work, we propose a transformer-based model for unsupervised representation learning of multivariate time series to predict the outcomes of soccer matches from leagues worldwide. this model is multitask in which it predicts the 1X2 probabilities and the exact scores (goals scored by home and away team).

***K*eywords** Transformers • Deep learning • Soccer

1 Introduction

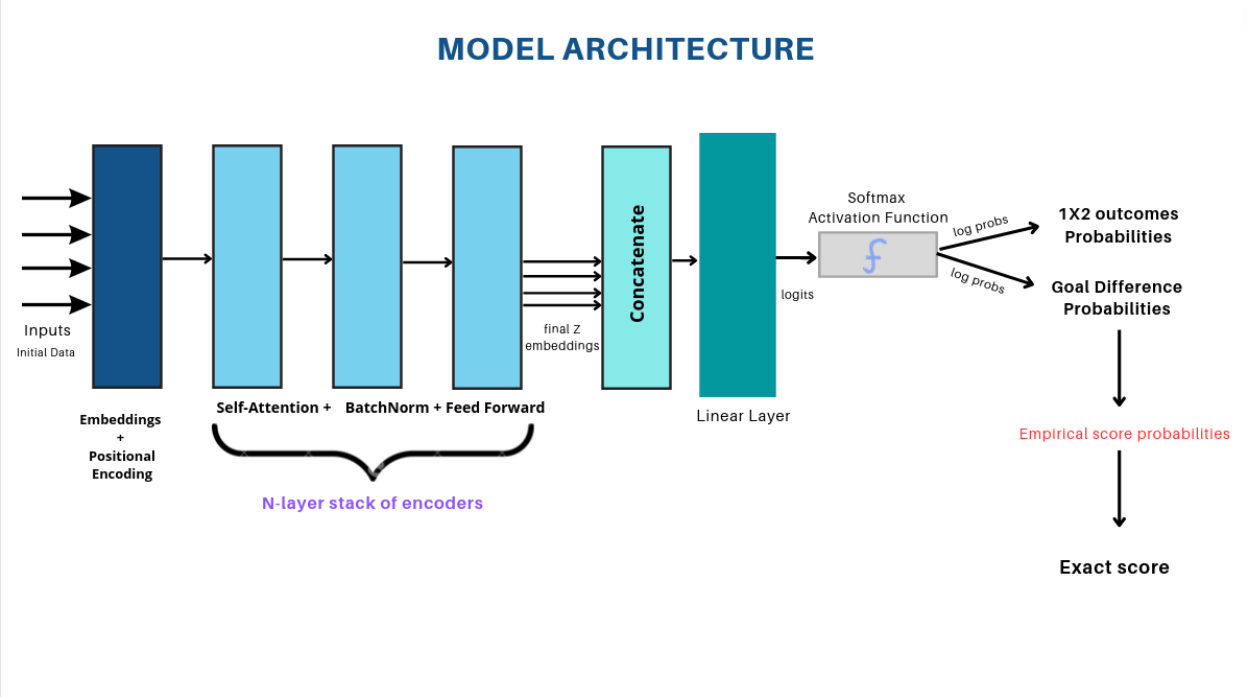
Soccer is the most popular sport in the world, and predicting the outcomes of matches has always been of great interest to sports enthusiasts, analysts, and betting companies. Traditional methods of prediction have relied on simple statistical models, but with the increasing availability of data and advancements in machine learning, more sophisticated models can be developed to improve prediction accuracy.

2 Model Architecture

In this work, we propose a `transformer-based`(Zerveas et al. (2021)) model for unsupervised representation learning of `multivariate time series` to predict `soccer` match outcomes from leagues around the world. Our model ingests team features inputs, each containing `statistical performance` realised during each played game. The model architecture consists of an `embedding layer`, a `positional encoder`, and an N-layer stack of `self-attention` and `feedforward` networks. The embedding output, the feedforward network's "Z" vectors, are concatenated to produce a single vector containing information on both `home` and `away` teams. This vector is subsequently transmitted through a `multitask` learning `linear layer`, which produces two sorts of results: `goal difference` and `1x2` outcome probabilities.

We evaluate the effectiveness of our model using `negative log likelihood` losses for both `1x2` and `goal difference` predictions to update the network parameters, and we can produce the `exact score`

of a particular game utilising goal differences probas and `empirical score` probabilities by competition.



Zerveas, George, Srideepika Jayaraman, Dhaval Patel, Anuradha Bhamidipaty, and Carsten Eickhoff. 2021. "A Transformer-Based Framework for Multivariate Time Series Representation Learning." In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery &Amp; Data Mining*, 2114–24. KDD '21. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3447548.3467401.