Public project summary

Product Recommender Systems in Sports Betting

Capstone Project Technical Discussion

Gyula Kiss

Master of Science in Business Analytics

Central European University

Department of Economics and Business

Budapest, Hungary

2021

Abstract

Personalized recommenders are apparent to anyone using web stores like Amazon or content providers like Netflix today. Based on implicit activity of users purchasing or watching products or explicitly rating items these e-commerce sites offer recommendations for other products to consider and help users to choose products from a large set.

The main distinction between recommender algorithms is based on the type of data that the algorithm uses: **content based** recommenders assume user profile and product content features to be identified as a preliminary feature engineering step. **Collaborative filtering** however use only users' ratings on products or transactional data of purchasing or interacting with products without identifying any features. The idea behind the latter is to predict the individual user's interest based on activities of many users.

Sports betting can be considered as selling bets to the players on various sport events. The product line in sports betting is very large, not just because of the abundance of sport events, but the various forms of bets (industry term is market), e.g. for a single football match players can bet on the winning team, number of goals, half time result, and dozens more. The significance of the project is to explore and investigate the matrix factorization machine learning methods and algorithms of recommenders and their adaptation to the Sports Betting industry, experimenting with real production data.

1 The Client and the industry

The client of this project is a global vendor of sports-betting platforms to lottery and gambling organizations across the globe. The Web retail channel provides abundant opportunities for personalization, and ease in data collection and a user interface that could be employed to recommend items in a non-intrusive way.

Key aspects:

- Feature engineering approach should to be avoided because that would be an extensive analysis project in itself
- Sports betting products (i.e. markets of sports events) have limited lifespan and ever-changing lifecycle, as matches are finishing in a limited time span and new matches are coming up continuously.

- In this highly non-stationary environment traditional machine learning techniques for model building are not feasible, we cannot train and build a model beforehand, because products age out from the system rapidly, we need an online learner solution
- The recommender system helps to filter products tailored for the player, but does not help to choose the outcome to bet on
- The recommender system's goal is to increase the player's betting activity without affecting the winning chances – the assumption is that a player is more likely to place a bet if the recommendation would match his preferences
- Courtesy of and with the help of sports betting operator partners, sufficient amount of real data is available to explore and validate the efficiency of the recommender system

2 The Data

Anonymized transactional data has been provided by a partner sports betting operator from archived data sources from the live production system for a period of 10 days in winter of 2021, a period when the leagues have been restarted after COVID and Christmas holidays.

The first decision we make is that we separate pre-match and live bets in our analysis, since the time frame of their availability is very different. Live bets are only available during the match, pre-match markets are published days or week before the match, but are available for betting before the match starts.

3 Method of model building and evaluation

Compared to standard collaborative filtering methods, we use recommendation by online learning: we process bets once and in the order they have occurred.

We replay each and every bet that occurred in the production bookmaker system and we do the following:

- Based on the actual global model and the specific player we predict the markets that the player is betting on with an ordered list, without knowing the bet that the player just placed. This is the top K product recommendation for the specific player at the specific time.
- Now we check the market that the player placed in this step, and compare with our recommendation list. We note the rank: what is the index of the real placed market on our predicted recommendation list. We also update and retrain our model according to the actual market that the player placed.

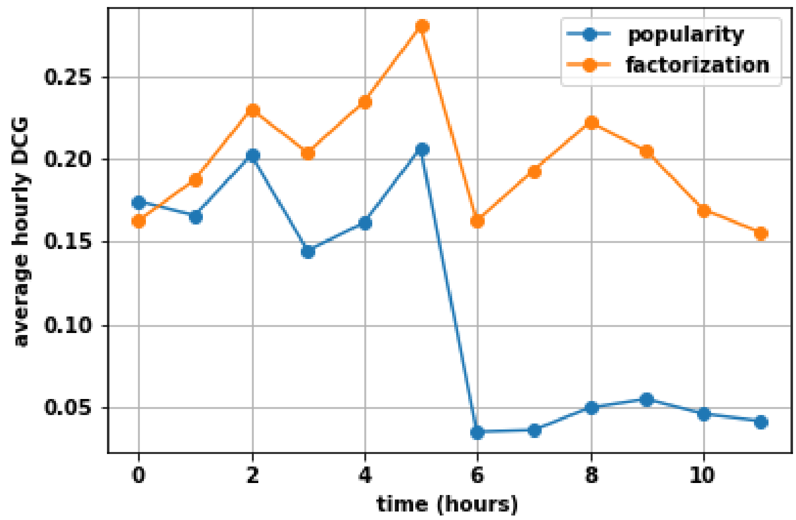
- We continue this for all bets and all markets in the bet.
- We calculate the online Discounted Cumulative Gain (DCG) score for each step for the specific timestamp of the bet and later we can average the DCG score for e.g. hourly periods to measure the performance of our model

4 Results

Our baseline model will be the obvious popularity model. We keep track of the number of the bets per markets, building a bestseller hit list. For players we will recommend the top K items from the bestseller list, regardless of the player, all player get the same recommendation in this model.

The online matrix factorization model can be fine-tuned by adjusting the hyper parameter. We inspect the dimension of factor matrices, negative rate, and Stochastic Gradient Descent algorithm's leaning rate.

We compare our models with the Discounted Cumulative Gain (DCG) score, a higher value means better recommendation. We see that the matrix factorization model performs better throughout the inspected period, when the popularity model starts performing poor, the matrix factorization model learns the situation quickly and gives good recommendations.



5 Conclusions

The result is that collaborative filtering approach with matrix factorization model gives surprisingly good results on our production data and gives us a feasible methodology that can be the basis to implementing a recommender feature in production.