УДК 004.855.5

ПРЕДСКАЗАНИЕ ELO-РЕЙТИНГА ИГРОКА НА ОСНОВЕ ШАХМАТНОЙ ПАРТИИ

Дудниченко М.Ю. (ИТМО)
Научный руководитель – Филатова А.А. (ИТМО)

Введение. Развитие шахматных программ и онлайн-шахмат позволили использовать машинное обучение для анализа партий. В данной работе рассмотрена задача предсказания ELO-рейтинга игроков на основе сыгранных ходов. Основными вопросами исследования являются: можно ли по ходам партии определить рейтинг игрока; если да, то какие признаки являются наиболее важными. Для создания датасета были использованы данные о партиях с сайта lichess.org за 2023 год. На основе ходов и компьютерного анализа партий предложены агрегированные признаки и построена модель для предсказания рейтинга.

Основная часть. Задача предсказания ELO-рейтинга игрока на основе ходов партии была рассмотрена ранее [1], однако исследование можно улучшить за счет использования более актуального датасета, добавления новых признаков и более глубокой интерпретации признаков. Также ранее были рассмотрены похожие задачи, а именно предсказать результат игры, используя только начало партии [2] и результат следующих партий на основе рейтинга игрока и других признаков [3]. Также отметим предсказание хода игрока в данной позиции с учетом его рейтинга [4].

Для исследования были использованы данные Lichess Database, а именно файл с играми за декабрь 2023 года, всего он содержит примерно 100 млн партий в формате PGN. Часть из них также содержат компьютерный анализ с оценками качества ходов и указанием ошибок, он является основой для создания признаков.

Целевой переменной является средний рейтинг двух игроков в партию, при этом была отдельно рассмотрена постановка задачи как регрессии и мульти-классовой классификации. Главные признаки основаны на качестве игры, а именно количестве ошибок и неточностей, потере преимущества, точности использования каждой фигуры, вероятности выиграть при данной позиции, сыгранном дебюте, времени на обдумывание хода и других.

Пайплайн состоит из следующих этапов: парсинг текстового файла PGN в табличный формат; отбор подходящих для анализа партий; создание признаков на основе ходов и их оценки компьютером; биннинг признаков; построение и интерпретация моделей. Так как важна интерпретируемость, были выбраны линейные модели и ансамбли деревьев решений. Лучший результат показала модель CatBoost, на ее основе был проведен анализ важности признаков. К самыми важным признакам партии модель отнесла название дебюта, среднюю потерю преимущества в ранней игре и за всю партию, а также процент равных позиций в партии. Биннинг признаков помог уменьшить переобучение модели.

Выводы. Получена модель для предсказания рейтинга игрока на основе сделанных ходов, также была проведена интерпретация результатов.

Список использованных источников:

1. Avva P., Hanke J. Guess the Elo – Predicting Chess Player Rating. – 2022.
2. Rosales Pulido H. A. Predicting the outcome of a chess game by statistical and machine learning techniques : дис. – Universitat Politècnica de Catalunya, 2016.
3. Thabtah F., Padmavathy A. J., Pritchard A. Chess results analysis using elo measure with machine learning //Journal of Information & Knowledge Management. – 2020. – Т. 19. – №. 02. – С. 2050006.
4. McIlroy-Young R. et al. Aligning superhuman ai with human behavior: Chess as a model system // Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. – 2020. – С. 1677-1687.