



AI Academy Dota2 Skill Prediction

AI-Academy - проект Сбербанка, направленный на популяризацию образования в сфере искусственного интеллекта среди школьников. В рамках проекта проходил хакатон, в рамках которого участникам предлагалось решить задачу оценки квалификации игрока в Dota 2 по данным его матчей. Материалы к [соревнованию по определению опытного игрока Dota2](#)

Здесь представлено моё решение, которое заняло 2 место на финальном этапе с точностью 0.91250.

Описание задачи

В игре Dota 2 участвуют две стороны: команды сил Света (The Radiant) и сил Тьмы (The Dire). Каждая команда состоит из 5 игроков — персонажей со своими уникальными способностями. Перед началом матча каждый игрок выбирает себе одного героя из 116 возможных. Цель игры — разрушить трон, находящийся на базе противника.

Повышая игровые навыки, игроки повышают свой рейтинг и получают медали опыта. Эксперты по игре Dota 2 утверждают, что легко могут отличить матч новичка от игры опытного игрока.

Задача — построить алгоритм, который сможет определять опытность игрока по статистике из одного матча.

Формат набора данных

Для решения задачи участникам предоставляется набор из почти 50000 примеров матчей, уровень игрока в которых известен. Для каждого из ≈ 21000 тестовых примеров, участникам необходимо проанализировать данные о матче при помощи своих алгоритмов и дать ответ — опытный ли был игрок. Допускается использование любых языков программирования, библиотек, методов машинного обучения, разрешено использование внешних данных из сети интернет.

Участникам предоставляются наборы данных с примерами матчей. Каждый пример описывает характеристики матча и статистику одного из игроков на момент окончания матча. Все примеры имеют уникальный идентификатор "id".

Всего два набора:

1. Обучающий (train), в котором для каждого примера известен тип медали опыта игрока.
2. Тестовый (test), для которых участникам неизвестна опытность игрока — ее необходимо вычислить.

Наборы данных предоставляются в двух форматах:

- CSV-таблицы: `skill_train.csv`, `skill_test.csv`
- JSON-документы: `skill_train.jsonlines`, `skill_test.jsonlines` (расширенный, содержит дополнительную информацию, которой нет в CSV-таблицах)

Описание полей с данными можно найти в руководстве по решению задачи. Про данные достаточно знать, что это некоторый набор чисел, идентификаторов и временных рядов, который потенциально несет в себе полезную информацию. Опыт игры в Dota 2 для использования данных не требуется.

Дополнительно предоставляются файлы с дополнительной информацией по героям, предметам и умениям.

Формат решения

Для каждого примера из тестового набора необходимо предсказать опытность игрока. В систему необходимо предоставить для проверки CSV-таблицу с предсказаниями, она должна содержать две колонки: id – идентификатор примера, skilled_prob - вероятность принадлежности игрока к классу 'Skilled'.

Оценка качества

Качество решения оценивается с помощью метрики ROC-AUC.

Материалы к задаче

- [материалы](#) Академии Искусственного Интеллекта Сббанка;
- разбор [похожих задач](#) на [тренировках по машинному обучению](#).

Для решения использован язык программирования Python, так как для него есть большое число библиотек для анализа данных: NumPy, Pandas, SciKit-Learn и другие. В качестве инструмента разработки – интерактивную среду Jupyter, в качестве основных моделей - CatBoost и LightGBM.

Базовый пример решения от организаторов доступен в виде Jupyter-тетради: [baseline/Dota2SkillPrediction_Tutorial.ipynb](#)