

Предсказание вероятности победы команды в Dota 2

Попов Артём

ВМК МГУ

Семинары по машинному обучению для 317-ой группы

Краткие сведения о Dota2

Dota 2 это:

- Dota 2 — многопользовательская компьютерная игра жанра MOBA.
- Две команды играют друг против друга
- В каждой команде по пять героев (герои не повторяются)
- Выигрывает тот, кто разрушит вражеское главное здание



Постановка задачи

Что дано:

- Обучающая выборка из матчей, для которых были известны некоторые их события и их результат.
- В тестовой выборке матчей история обрезана на первых 5 минутах.

Что необходимо:

- Предсказать для тестовых матчей вероятность победы команды Radiant.

Что можно:

- Использовать любые дополнительные данные.

Главное команды, а не игроки

Основная идея при преобразовании признаков

- Стараться сделать БОЛЬШЕ признаков, характеризующих КОМАНДУ
- Стараться сделать МЕНЬШЕ признаков, характеризующих ОТДЕЛЬНЫХ ИГРОКОВ
- One-hot-encoding героев:
 - +1, если герой играет за radiant
 - -1, если герой играет за dire

Одна из первых удачных попыток:

- Убрать все индивидуальные показатели героев, заменив их на среднее (и т.п.)
- Было хорошо, но был способ получше

Преобразование исходной обучающей выборки

Все действия над исходной обучающей выборкой (отсортировано в порядке полезности):

- One-hot-encoding героев
- Сортировка индивидуальных признаков героев (опыт, накопленное золото и т.д.) внутри каждой команды
- Добавление сумм индивидуальных признаков героев внутри каждой команды
- Добавление отношения смертей к убийствам
- One-hot-encoding lobby type
- Учитывать смерти в степени 1.1
- Замена пропущенных значений на медианы

Добавление информации из сырых матчей

Из матчей можно было извлечь информацию о предметах, купленных героями, и выбранных каждым героем способностей.

Мешок из предметов

- Много признаков
- + Интерпретируемо
- + Даёт выигрыш при использовании и бустинга и лог. регрессии

Мешок из способностей

- Очень много признаков
- Неинтерпретируемо
- + Даёт выигрыш при использовании бустинга

Дополнительные признаки

- Синергия и антисинергия
 - Предложенная на форуме идея, как учитывать парное взаимодействие героев.
 - Как и на форуме, я использовал два признака.
 - Если применять в лоб, то ведёт к переобучению.
Необходим подход, как при работе с счётчиками.
Выбранное число фолдов — 30.
- Характеристики героев
 - Дополнительная информация с официального сайта Dota2
 - Распределение героев по ролям в игре, тип их атаки, принадлежность характеристике
 - Можно собрать статистику по каждой команде
 - Даёт совсем небольшой прирост

Используемые алгоритмы

- Логистическая регрессия (Skicit-learn)
 - Ко всему, кроме ролей героев и выбранных способностей.
 - Чем больше регуляризация тем хуже.
- Градиентный бустинг (XGBoost)
 - Ко всему, что было показано.
- Итоговая модель:

$$ALG = 0.65LR + 0.35GB$$

Результат

Результаты в разных лидербордах:

- Private: 0.76455
- Public: 0.76490
- All: 0.76474 (большой контекст)

Простое наблюдение: хорошо на одной половине теста, хорошо на всём тесте \Rightarrow хуже на другой половине теста