

Решение конкурса

Dota 2: Win Probability Prediction (3 место)

Каюмов Эмиль

ММП ВМК МГУ

Семинар «Машинное обучение»

25 марта 2016

Задача

- Задача: предсказать вероятность победы команды radiant по информации о первых пяти минутах матча Dota 2.
- Имеется JSON с информацией о ходе игры и набор избранных признаков в виде таблицы для удобства.
- Метрика: AUC.

План

1 Модель 1

- Работа с базовыми признаками
- Метапризнаки по мешкам слов
- Обучение модели

2 Модель 2

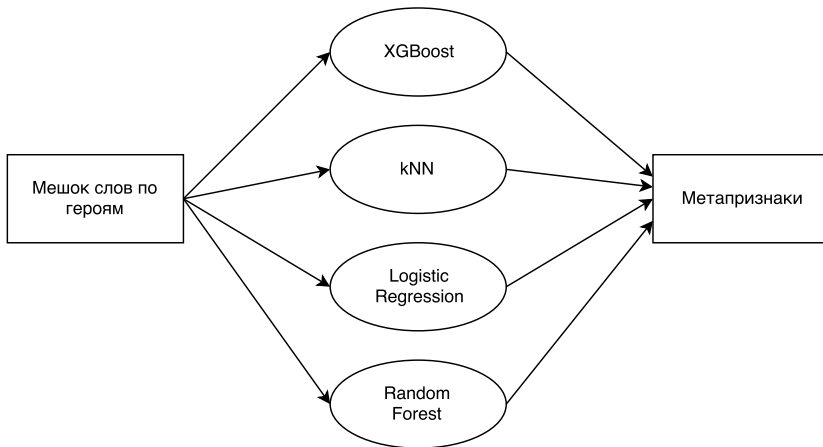
- Использование дополнительной информации
- Метапризнаки
- Синергия
- Обучение модели

3 Финальная модель

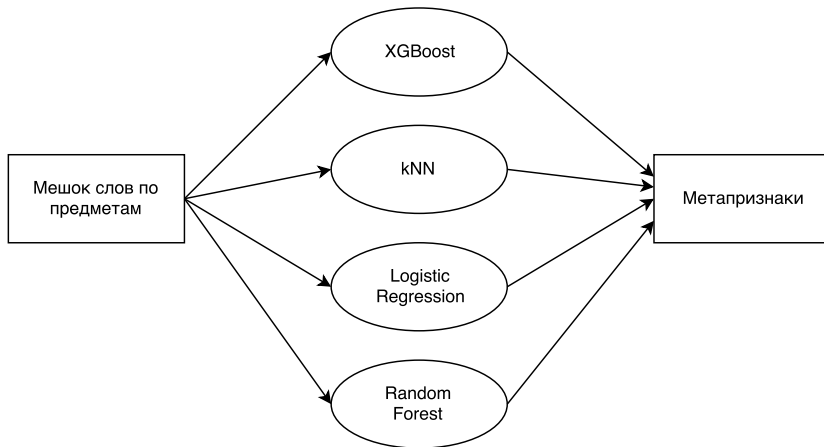
Работа с базовыми признаками

- ❶ Посчитаем характеристики, характеризующие команды:
 - Разности сумм показателей команд ($\sum_{i=1}^5 r_i - \sum_{i=1}^5 d_i$)
 - Отношение сумм показателей команд ($\frac{\sum_{i=1}^5 r_i}{\sum_{i=1}^5 d_i}$)
 - Разности максимальных/минимальных значений показателей в команде ($\max(r_i) - \max(d_i)$, $\min(r_i) - \min(d_i)$)
 - Разности среднеквадратичных отклонений показателей в команде.
- ❷ Добавим признаки: максимальные значения xp по командам, разность времени получения bottle.
- ❸ Удалим индивидуальные показатели (кроме xp) и некоторые общие показатели (lobby_type, ward_sentry_count, ward_observer, first_blood, ...)

Метапризнаки по мешку слов из героев



Метапризнаки по мешку слов из предметов



Обучение модели

Используем Vowpal Wabbit.

- Обычные признаки, после описанных преобразований.
- Метапризнаки ($4 + 4$).
- Мешки слов по героям, предметам и умениям.

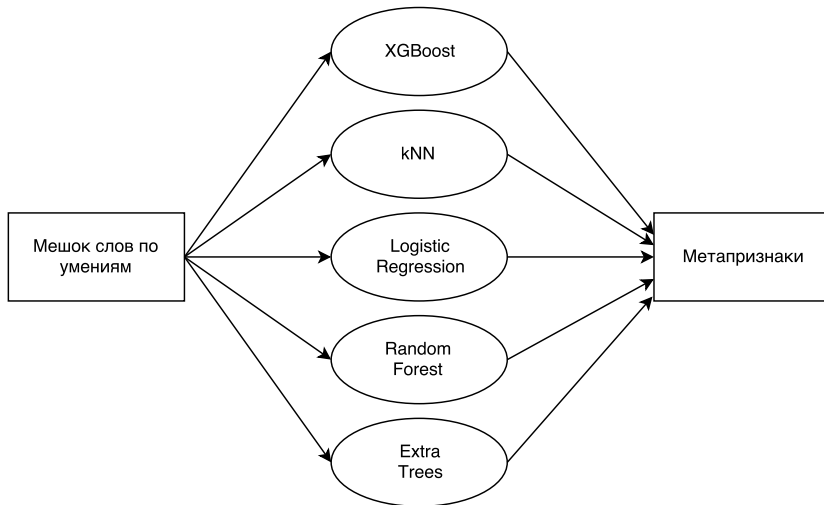
Private LB: 0.76413 (лучшая одиночная модель).

Использование характеристик героев

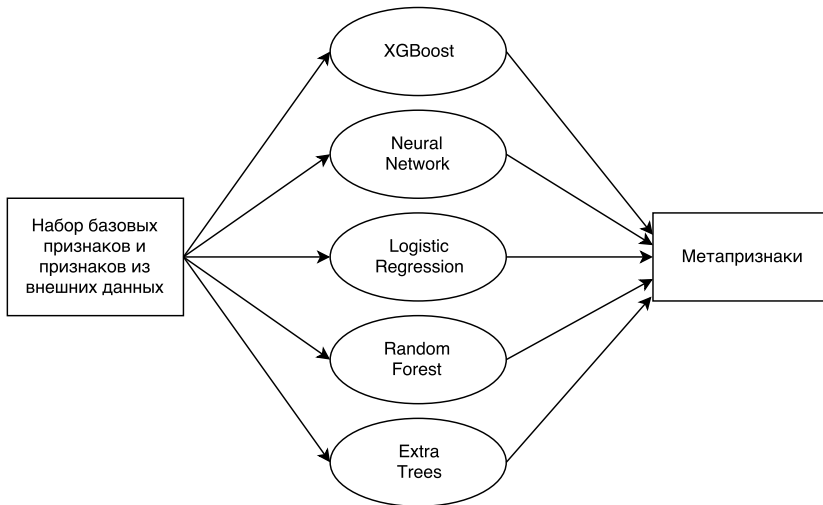
Для каждого из героев существуют такие сведения, как strength, intelligence, agility, speed, range, attacktime, и их изменения с ростом уровня героя.

- Добавим признаки, характеризующие команду целиком по новым данным (аналогично работе с базовыми признаками).
- Посчитаем, сколько раз каждая из ролей встречается в команде. Возьмём разность для двух команд.
- Также удалим часть индивидуальных характеристик героев.

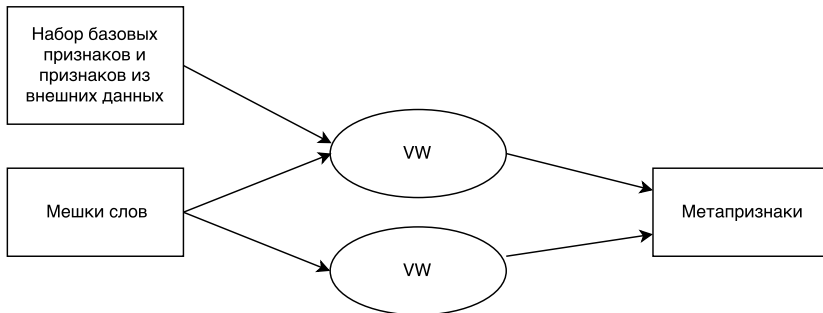
Метапризнаки по мешку слов из умений



Метапризнаки по обычным признакам



Метапризнаки с помощью Vowpal Wabbit



Синергия

Как указывалось на форуме, можно учитывать связки героев без перехода к квадратичным признакам.

- 1 Посчитаем по обучающей выборке, сколько раз каждая пара героев приводила к победе и поражению команды. Кроме того, посчитаем пары героев противоположных команд, чтобы учесть «доминирование» одного героя над другим.
- 2 Нормализуем эти показатели, чтобы получить доли.
- 3 Посчитаем разности сумм этих показателей по героям матча. Получим агрегированные показатели по командам.
- 4 Возникает переобучение аналогично счётчикам для категориальных признаков – необходимо использовать фолдинг.

Обучение модели

Используем Vowpal Wabbit.

- Обычные признаки из предыдущей части.
- Признаки из внешней информации.
- Все метапризнаки ($4 + 4 + 5 + 7$).
- Синергия и антисинергия.
- Мешки слов по героям, предметам и умениям.

Трудно настраивать из-за метапризнаков и синергии.

Private LB: 0.76266 (хуже предыдущей модели).

Финальная модель

- Час до конца контеста.
- Всего один оставшийся сабмит.
- Два обученных Vowpal Wabbit.
- Валидация не работает.

Пусть: $new = 0.8 \times VW_1 + 0.2 \times VW_2$

Private LB: 0.76458 (против 0.76413 и 0.76266)