

Исследование Лиги легенд

Шандыбин Андрей

May 2020

1 Вступление

В современном мире компьютерные игры стали частью нашей жизни. Например, такие как League of Legends. Смысл которых в противостоянии 2 команд, где ничьи не может быть. Я увлекаюсь одной из таких игр и мне стало интересно с какой максимальной точностью можно предсказывать победу одной из команд используя методы машинного обучения. Так как может быть только 2 исхода игры, то эту проблему представить как задачу бинарной классификации. В данном исследовании выбрана наиболее подходящая модель и, перебирая гиперпараметры для нее, нашел наилучшие на которых достигается максимальное качество. Также если собрать больше статистики на всех промежутках игры, то может быть создан показатель, который показывает шанс на победу одной и другой команды. Это поможет лучше анализировать игру, людям незнакомым с игрой позволит лучше её понимать.

2 Постановка задачи

Каждая команда состоит из 5 игроков, каждый играет на свое уникальном герое. Цель игры разрушить главное здание противника, в процессе игроки добивают миньонов и монстров, за что получают опыт и золото, на что могут покупать предметы, чтобы стать сильнее. Надо обучить модель, чтобы по входным данным, модель предсказывала победу какая команда победит. (?)

3 Датасет

Датасет, на котором было проведено исследование, состоит из 19 статистических характеристик для каждой из команд за первые 10 минут, в сумме 38 признаков, когда вся игра длится от примерно 30 минут. Также содержит информацию о победе одной из команд, она представлена единицей в случае победы, нулем в случае поражения соответственно. Также можно улучшить модель добавив туда данные которые содержатся в данных

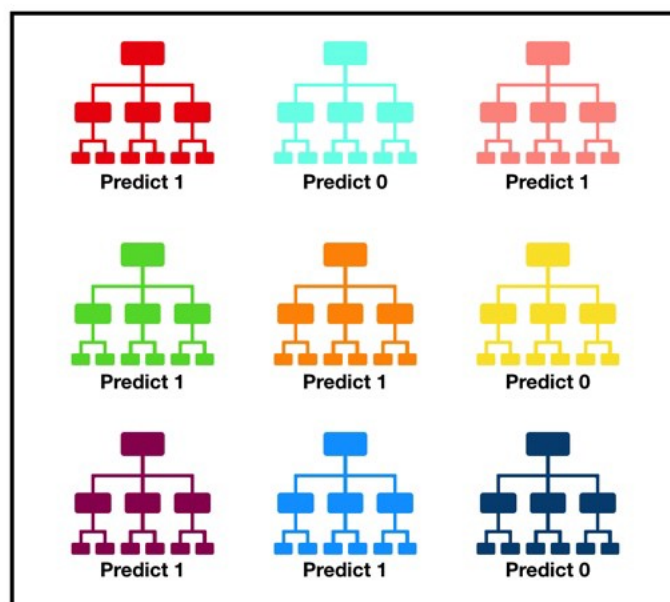
только в косвенном виде. Так было серьезно увеличено качество модели, добавив признак отвечающий за процент нейтральных миньонов из всех добытых каждой из команд.

4 Описание методов

Для начала предстояло выбрать модель, мой выбор предлагается решающий лес, так как он заслуженно является одним из лучших для решения подобных задач.

Решающий лес это комбинация отдельных, несвязанных, что очень важно, между собой решающих деревьев. А каждое дерево имеет следующий принцип работы. Дерево разделяет на каждом этапе своей работы данные по некоторому признаку, который позволит их разделить так, чтобы результирующие группы были настолько отличны друг от друга, насколько это возможно (и члены каждой результирующей подгруппы были настолько похожи друг на друга, насколько это возможно)

Случайный лес как уже было указано, состоит из большого количества независимых деревьев, которые работают в группе. Каждое дерево предсказывает класс и потом класс, который предсказало большинство является результатом предсказания все модели.



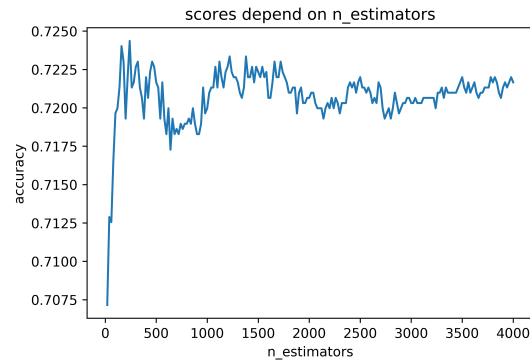
Tally: Six 1s and Three 0s
Prediction: 1

Рис. 1: Random Forest

Такая модель лучше так большое количество относительно несвязанных моделей (деревьев), действующих вместе, превзойдет любую из отдельных составляющих моделей. Потому что деревья защищают друга друга от ошибок, когда какие-то могут быть неправильными, в это время много других будут правильными.

5 Эксперименты

Для начала стоило найти количество деревьев в модели. Для всех измерений была обучена модель и результат был усреднен по 10 предсказаний, чтобы минимизировать влияние случайности. Сначала был сделан перебор всех моделей начиная от 20 деревьев в лесу дальше надо выбрать оптимальное максимальное количество деревьев. Так как лес не переобучается, поэтому можно пробовать достаточно большие значения, но с некоторого момента результаты будут примерно одинаковыми, в данном случае стоит выбрать 4000, как максимальное количество деревьев в лесу.



Затем выбрав количество деревьев, где достигается максимальное качество, начинается перебор остальных гиперпараметров классификатора в наборах, такие как максимальная максимальная глубина деревьев, функция оценки качества разделения, минимальное количество элементов, которое должно быть в листе или узле и так далее. Также усредняя результат по 10 предсказаниям. Гиперпараметры надо перебирать именно в наборах, потому некоторые зависят друг от друга. При этом получен вот такой график при :

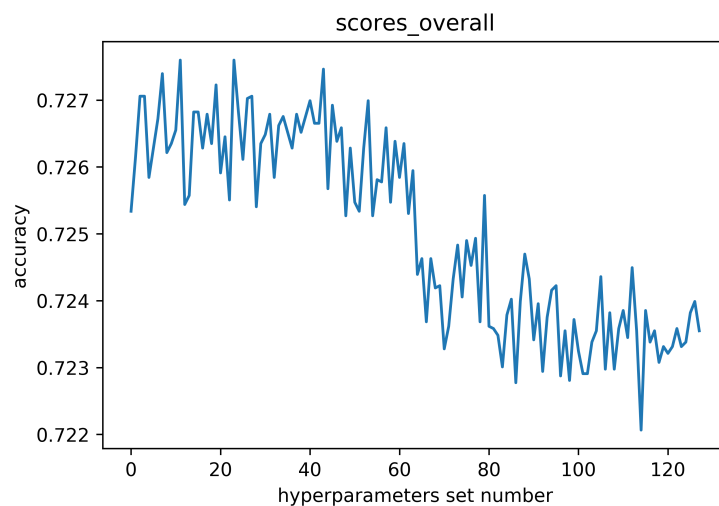


Рис. 3: sets of hyperparameters

Выбрав лучшие гиперпараметры, после добавления нового признака в имеющийся датасет, были получены очень хорошие результаты. Увеличив максимально качество с 0.7273, до 0.7321

6 Выводы и заключение

Таким образом с помощью случайного леса, получена модель, которая достигает очень хорошего качества. И может по данным всего лишь за первые 10 минут игры, которая в среднем длится 30 минут, предсказывать какая из команд победит с вероятностью равной 73.2%. Это позволит создать такую же модель для каждого момента игры, что позволит лучше анализировать геймплей Лиги Легенд.