

Слайд 1

Здравствуйте уважаемые члены комиссии. Я Перехожих Алексей Дмитриевич, студент группы 191-363. Тема моей работы «Прогнозирование результатов киберспортивных матчей в игре «Rocket league» с применением технологий больших данных

Слайд 2

Актуальность данной темы обусловлена активным развитием и популярностью компьютерного спорта, а также отсутствием доступных решений для прогнозирования матчей. Целью работы было выбрано прогнозирование результата предстоящих киберспортивных матчей в игре «Rocket league», используя статистические данные прошедших игр.

Слайд 3

Как и классические виды спорта, компьютерный спорт имеет свою зрительскую аудиторию. По нему проводятся крупные соревнования с большими призовыми, а крупные компании спонсируют различные организации, лиги и команды.

Существует множество киберспортивных игр, но в данной работе рассматривается «Rocket league» – игра в жанре спортивного симулятора, где соревнуются две команды, ее суть схожа с футболом, надо забить в ворота соперника мяч и защищать свои, только игроки управляют не людьми, а радиоуправляемыми машинками.

Слайд 4

Одна из основных причин выбора данной предметной области для работы - это удобство записи и доступность данных киберспортивных матчей. Записи матчей содержатся в «Реплеях». В одном таком файле хранится набор различной информации о матче, считывая которую инструментами повтора игры, можно полностью воспроизвести матч. Такой формат позволяет значительно экономить место, а также позволяет использовать информацию из игр для создания интересных проектов.

Слайд 5

Примерами таких работ являются проекты на основе больших данных, такие как определение xG(показателя ожидаемых голов в матче), предоставление расширенной статистики по обработанным данным реплея, проекты, используемые для маркетинга. В ВКР информация из реплеев используется для прогнозирования предстоящих матчей.

Слайд 6

В математическом смысле данная работа представляет собой задачу классификации, где по входным данным определяется принадлежность

результата одному из двух классов (победа или поражение команды). В исходных данных существует конечное множество объектов (матчей), с различными параметрами, а также доподлинно известно, какому классу они принадлежат.

Поскольку на высоком уровне множество параметров из расширенной статистики являются лишь отражением стиля игры команд и слабо коррелируют с результатом, то было решено использовать рейтинговые модели. Первая модель – ELO. В ее основе лежит идея присвоения каждому игроку числового значения, отражающее его навыки и силу относительно других игроков. По мере свершения игр рейтинг изменяется в зависимости от результатов матча и ожидаемого исхода.

Слайд 7

Модель Gliko по своему смыслу схожа с предыдущей, но имеет дополнительный параметр «Достоверности рейтинга» и более сложные расчеты изменений. При появлении нового игрока мы слабо уверены в его рейтинге, а потому разброс его возможностей больше, как и изменения рейтинга после каждого его матча. Также эта система позволяет учитывать те моменты, когда игрок не выступает долгое время, в таких случаях уверенность в его рейтинге падает с каждым игровым периодом.

Слайд 8

Так как основной режим игры это 3 на 3, при том что в нем нет четкого распределения игроков по ролям, то в работе применяется распределение полученных или потерянных очков рейтинга между игроками, в зависимости от их вклада в результат матча и ожидаемых значений

Слайд 9

Для реализации моделей использовался язык программирования Python и некоторые его модули, а также Excel таблицы. Для получения информации использовались технологии веб-скрейпинга, парсинга и API, для хранения информации таблицы ключ-значение, реализованные в виде словарей в Питоне.

Слайд 10

Информация о загруженных судьей реплеях собиралась с сайта в виде JSON файлов, а затем нужные данные обрабатывались в Excel таблицу.

Слайд 11

На слайде перечислены функции программы, написанной для реализации моделей. Структура изначально выстраивалась таким образом, чтобы работать с разными рейтинговыми моделями, меняя лишь используемые функции.

Слайд 12

После нахождения экспериментальным путём оптимальных изменяемых параметров, было выделено три наилучших модели. После сравнения была выделена модель Gliko с параметрами эпохи=1 день, $C_{RD} = 36$ и $\text{Gamma}=2$, она используется для всех дальнейших прогнозов.

Слайд 13

Первый турнир, не входящий в исходные данные – NA Spring cup, проходил в конце мая. Он состоит из сетки игр на вылет со вторым шансом (называемым нижней сеткой). До начала соревнований был сделан прогноз без использования модели, и с ее использованием. Программа смогла верно определить победителя и финальную пару, и тем самым показала себя значительно лучше прогноза эксперта. При моделировании по раундам, модель правильно определила исход 26 из 30 матчей. Причиной показателя точности значительно выше ожидаемой может быть фактор того, что на турнире играли известные команды, о которых уже имеется достаточное количество информации в датасете и такие игры предсказываются лучше, чем игры между двумя малоизвестными командами на более ранних стадиях турниров.

Слайд 14

Так как модель хорошо себя зарекомендовала в прогнозировании крупных турниров, было решено сделать прогноз на Европейский турнир, проходящий в середине Июня в том же формате, что и описанный ранее северо-Американский. Основываясь на образованный рейтинг команд, модель определила «Karmine Corp» победителем. Однако, промоделировав турнир сто тысяч раз, основываясь на выдаваемых системой вероятностях победы, был получен результат, что в 32.4% случаях должны победить «Vitality». Вероятнее всего это связано с более удачной сеткой для команды и небольшой разницей в рейтинге.

Практическая значимость полученного результата, заключается в ценности прогнозируемой информации и написанной программы для фанатов игры, а также для киберспортивных аналитиков, букмекеров и организаторов турниров.

Слайд 15

Спасибо за внимание. С удовольствием отвечу на Ваши вопросы.