

Министерство образования Республики Беларусь  
Учреждение образования  
Белорусский государственный университет  
информатики и радиоэлектроники

УДК 004.85

Михолап  
Алесь Александрович

Ранжирование игроков в командных видах спорта

**АВТОРЕФЕРАТ**

на соискание степени магистра информатики и вычислительной техники  
по специальности 1-40 81 01 — Информатика и технологии разработки  
программного обеспечения

Научный руководитель  
Сержанов Максим Валерьевич  
кандидат технических наук, доцент

Минск 2016

# ВВЕДЕНИЕ

## Общие сведения

Во все времена люди задавались вопросом определения лучших в какой-либо сфере деятельности. Для его разрешения проводились многочисленные турниры, состязания, голосования. Он стал еще более острым в современном глобальном обществе, в котором как никогда сильны соперничество и конкуренция. Каждый из нас постоянно оценивает различные варианты: за какую спортивную команду болеть, как провести выходные, куда пойти за покупками. Интернет заполнен топ-10 списками на все возможные темы, спортивные газеты совместно с экспертами регулярно ранжируют команды по их результативности, а в конце года традиционно подводятся итоги и определяются самые значимые события, влиятельные люди и так далее.

В большинстве случаев суждения такого рода делаются людьми субъективно на основании разных критериев. Данный подход достаточно хорошо удовлетворяет повседневные потребности, но его нестабильность может оказаться критическим недостатком, когда на основании этих суждений принимаются важные решения. Например, какие ценные бумаги приобретет инвестиционная компания, или какие ставки даст своим клиентам букмекер. В таких вопросах обычно руководствуются мнением экспертов, но даже они могут не учесть всего многообразия обстоятельств или наоборот быть подвержены ложному впечатлению. С целью решения данной проблемы осуществляется формальная постановка задачи, выявление значимых факторов и моделирование их взаимодействия для получения объективного ответа. Если даже без экспертного контроля не обойтись, алгоритмическое решение предоставляет неплохую отправную точку.

В этой работе фокус делается на исследовании предсказательной способности систем рейтингов спортивных состязаний. Эта проблема занимает центральное место в задачах составления списков лучших игроков и команд и прогнозирования исходов будущих матчей. В обоих случаях ее решением заинтересованы крупные коммерческие индустрии — спортивные СМИ и букмекерские конторы.

## Постановка задачи

Основой всех обобщенных систем рейтингов являются исходы прошедших матчей.

Специфичные модели учитывают множество особенностей конкретного вида спорта, например количество пасов в футболе или подборов в баскетболе. Они могут пытаться уловить тонкие особенности каждого из участников

состязаний, чтобы детально симитировать возможные варианты развития событий. Однако, такие модели требуют значительно большего количества данных и неприменимы к иным видам спорта. Более того, в открытом доступе не удалось найти сведения об их успешном применении в задаче ранжирования. Возможно, они полезны в других задачах, например, для обнаружения перспективных игроков при драфте по их статистическим показателям.

Обобщенные модели используют бинарные исходы (победа/поражение) или финальный счет встречи (Margin of Victory). Это позволяет использовать одну модель для различных видов спорта, меняя некоторые настраиваемые параметры, но сохраняя общую структуру. Очевидный недостаток — игнорирование всей дополнительной информации. Однако, на практике этот недостаток не столь критичен, так как подготовить набор данных, содержащий эту информацию достаточно сложно, а порой и невозможно, например, когда не сохранилось даже видеозаписи игры и известен лишь ее результат. Поэтому, в данной работе будут использоваться только обобщенные модели.

Формально историю игр можно рассматривать как упорядоченный по дате набор матчей  $m_1, \dots, m_N$ , где каждый матч представлен парой команд  $m_k = (t_i, t_j)$ , а команда — набором игроков  $t_k = (p_1, \dots, p_M)$  и известен исход матча  $o_k$ .

Упорядоченность по дате естественна и может использоваться как важный фактор модели, например, ближайшие матчи учитываются с большим весом, чем давние.

В данной работе рассматриваются только парные состязания. Некоторые системы способны учитывать более двух участников встречи, но на практике подобные виды спорта составляют меньшинство как по количеству, так и по популярности.

Команда состоит из  $M$  участников, где  $M$  зависит от типа игры, например, одиночный или парный теннис.

Исход матча  $o_k$  может быть победой первой команды, победой второй команды, либо ничьей.

# ОБЩАЯ ХАРАКТЕРИСТИКА

## Актуальность темы исследования

Системы ранжирования игроков и команд могут применяться для решения многих практических задач. В их число входят поиск перспективных игроков из молодежной лиги, определение лучших команд по результатам прошедшего сезона и прогнозирование исходов будущих матчей. Каждая из этих задач, как правило, решается на основании экспертной оценки. Однако, в сфере профессионального и любительского спорта она не всегда точна и зачастую содержит субъективные предпочтения человека. Кроме того, даже профессиональным аналитикам сложно учесть все прошедшие матчи и транзитивные отношения между командами. Математическая модель лишена этих недостатков и способна дать более стабильный и, зачастую, точный прогноз.

## Степень разработанности проблемы

Простейшие системы рейтингов существуют уже долгое время, но сравнительно сложные и гибкие модели появились не так давно. У алгоритма, предложенного в данной работе есть две основных альтернативы — Elo и TrueSkill.

Elo широко известна среди шахматистов, так как именно она применяется для определения ранга каждого профессионального игрока. Ее использование FIDE началось в 1970 году и продолжается до сих пор. Это свидетельствует о высокой степени коррелированности определенного ей рейтинга и истинного уровня мастерства. Со временем Elo приобрела популярность и в других видах спорта. Ее сильными сторонами являются относительная простота и интерпретируемость, а также стабильность и устойчивость к неожиданным случайным событиям.

TrueSkill является логическим продолжением Elo, разработанным в компании Microsoft с целью определения рейтингов игроков для автоматического составления команд в компьютерной игре Halo 2. Она значительно сложнее своего предшественника, но позволяет более тонко учитывать динамику изменения навыков игрока. Ее подход основывается на вероятностной постановке задачи, что открывает перспективы учета множества различных факторов, но одновременно увеличивает вычислительную и алгоритмическую сложность всей системы.

Эти методы достаточно хорошо справляются со своей задачей, но они были спроектированы для ранжирования игроков. Предлагаемый алгоритм нацелен на предсказание исходов будущих матчей, что является несколько иной проблемой, которая не столь пристально рассматривалась альтернативными системами.

## **Цель и задачи исследования**

Целью исследования ставилась разработка новой модели системы рейтингов, обладающей большей предсказательной силой, чем существующие альтернативы. Для ее достижения был проведен анализ предметной области, поиск литературы, выработка подхода постановки задачи. Это позволило определить возможные пути ее решения и построить в соответствии с ними теоретическую математическую модель. Следующей задачей была реализация программного продукта для проведения экспериментов и практического применения модели.

## **Теоретическая основа исследования**

В данной работе повсеместно используется предположение о вероятностном характере данных. Вероятностный подход обладает адаптивностью, необходимой для использования математической модели в реальных условиях, когда окружающая среда подвержена одновременному воздействию множества факторов. Учесть их все практически невозможно, поэтому их кумулятивный эффект считается случайной величиной с определенным распределением и параметрами. Теоретическая база такого рода методов хорошо проработана, в частности, были использованы результаты работ Арпада Эло, Томаса Минки, Лео Бреймана.

## **Основные положения, выносимые на защиту**

а) Необходимость применения вероятностного подхода при построении систем рейтингов. Показано, что он обладает значительными преимуществами в сравнении с предыдущими алгебраическими методами.

б) Невозможность оценить значение рейтинга игрока в ближайшем будущем по динамике его изменения в прошлом. Показано, что временной ряд значений рейтингов не коррелирован с любыми своими сдвигами.

в) Комбинирование базовых моделей, построенных определенным образом, приводит к значительному улучшению точности предсказаний. Описаны экспериментальные результаты применения комбинации к реальным спортивным матчам и проведено их сравнение с результатами базовых моделей.

## СТРУКТУРА РАБОТЫ

**Введение** содержит описание проблемы, мотивацию для ее решения и общие сведения о предметной области. Там же осуществляется постановка задачи и общая характеристика применяемого класса моделей.

**Глава 1** описывает существующие подходы с указанием их достоинств и недостатков. Она структурирована таким образом, чтобы показать историю развития систем рейтингов от простых и интуитивных к сложным и точным. Рассматриваются как алгебраические методы, основанные на свойствах матриц и графов, так и вероятностные методы, оперирующие случайными величинами.

**Глава 2** содержит информацию об использованных наборах данных. Все этапы исследования были провалидированы на трех различных видах спорта: футбол, Dota 2 и League of Legends. В данной главе раскрываются технические и концептуальные особенности каждого из них.

**Глава 3** описывает ход эксперимента по моделированию рейтингов игроков временными рядами и применению известных методов прогнозирования для оценки будущих рейтингов. Приводится мотивация применять моделирование временными рядами, процесс предварительной обработки данных, детали использованных методов и результаты эксперимента.

**Глава 4** является основным разделом работы, содержащим описание алгоритма композиции моделей. Сравниваются альтернативные подходы агрегирования базовых моделей с перечислением сильных и слабых сторон каждого из них. Дается подробная информация о проведении эксперимента, включая критерии качества, результаты на реальных данных и особенности использованного метода.

**Глава 5** содержит описание программного комплекса, реализованного для целей данной работы. Описывается техническая база исследований, использованные внешние модули, и детальная архитектура системы.

**Заключение** подводит итоги работы, включая достигнутые результаты и перспективы дальнейших исследований.

## **ЗАКЛЮЧЕНИЕ**

В результате выполнения работы были проанализированы существующие системы рейтингов и предложен способ их комбинирования, позволяющий добиться существенного улучшения точности прогнозов. Универсальность предложенного метода позволяет применять его к широкому классу базовых систем без необходимости учитывать специфику работы каждой из них.

Он, в отличие от альтернативных подходов, направлен на максимизацию предсказательной способности, что позволяет эффективно использовать его для оценки вероятностей исходов будущих спортивных матчей.

Для демонстрации практических результатов было разработан программный комплекс, содержащий весь необходимый набор функций. С его помощью можно получать данные из внешних источников, преобразовывать и хранить их в удобном формате и осуществлять все описанные в данной работе алгоритмические процедуры.

Итоги исследований были представлены и опубликованы на 52-й научно-технической конференции аспирантов, магистрантов и студентов БГУИР 2016.

Дальнейшее направление работы может включать поиск альтернативных систем рейтингов для включения в комбинацию в качестве новых базовых компонент.

## **СПИСОК ОПУБЛИКОВАННЫХ РАБОТ**

1 - А. Михолап А.А. Ранжирование игроков в командных видах спорта / А.А. Михолап // 52-я научно-техническая конференция аспирантов, магистрантов и студентов БГУИР: Тезисы доклада - Минск, 2016 - С. 33 - 34.

Библиотека БГУИР