

ПРИМЕНЕНИЕ МЕТОДОВ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ ДЛЯ ПРОГНОЗИРОВАНИЯ РЕЗУЛЬТАТОВ СПОРТИВНЫХ СОБЫТИЙ

Бабийчук Г.А.¹, Ляпунов С.М.²

¹Бабийчук Георгий Александрович – бакалавр, студент;

²Ляпунов Сергей Михайлович – бакалавр, студент,
кафедра инструментального и прикладного программного обеспечения,
Институт информационных технологий
Московский технологический университет,
г. Москва

Аннотация: в данной статье рассматривается применение методов интеллектуального анализа данных в спорте. Основное внимание уделено области прогнозирования результатов матчей и селекции игроков. Были рассмотрены примеры использования методов анализа данных для прогнозирования хоккейных и футбольных матчей. Также приведено описание работы по исследованию данных баскетбольной статистики матчей Национальной баскетбольной ассоциации.

Ключевые слова: интеллектуальный анализ данных, спортивная статистика, OLAP, баскетбол, NBA.

Анализ данных используется для извлечения нетривиальной и потенциально полезной информации из огромных наборов данных, доступных в различных областях науки и человеческой деятельности.

Современные методы интеллектуального анализа используются для решения проблем ассоциации, классификации, диагностики и прогнозирования во многих областях человеческой деятельности, в том числе и в такой области как спорт.

В последнее время отмечается нарастающее использование методов интеллектуального анализа данных для поддержки принятия решений по всем аспектам профессионального спорта.

Так, например, в рамках улучшения тренировочного процесса была предложена базированная на кинезиологии экспертная система, которая идентифицирует неадекватные физиологические и психологические параметры спортсменов и предупреждает о них.

Метод главных компонент, методы линейной регрессии и Байесовское моделирование получили свое применение в одной из распространенных проблем в профессиональном спорте – травмах. С целью прогнозирования и предотвращения травм был проведен проект по анализу множества факторов подготовки игроков спортивной команды [1]. В исследовании участвовало 20 игроков команды на протяжении одного сезона. Рассмотрено более 2000 измерений с 150 параметрами антропометрических, физиологических характеристик, тренировочной нагрузки и характера травм. Результаты анализа могут быть использованы для улучшения управления процессом подготовки команды.

Интеллектуальный анализ данных также нашел применение при анализе спортивного видео, идентификации спортивных талантов, управлении клубами, однако отдельное внимание в рамках данной работы хотелось бы уделить прогнозированию результатов матчей и селекции игроков.

Для максимально достоверного прогноза результатов матчей командных видов спорта необходимо проведение анализа большого набора данных спортивной статистики. Подобные статистические данные содержат информацию не только о команде в целом, но и о действиях каждого игрока в отдельности, а, следовательно, селекция игроков является важной задачей для конечного успеха команды.

Процесс отбора игроков и формирования команды — мультикритериальная проблема принятия решений, где конечный успех определяется тем, насколько набор отдельных игроков образует эффективную команду.

Впервые анализ статистических данных игроков был успешно применен в профессиональном спорте в 2002 году генеральным менеджером бейсбольной команды «Oakland Athletics», который, имея одну из самых маленьких платежных ведомостей, смог создать конкурентно способную команду, вышедшую в плей-офф. Успех «Oakland Athletics» побудил многие команды, включая другие профессиональные спортивные лиги, начать собственные исследования в данной области. В итоге, последние работы опираются уже не на простой статистический анализ, а проводятся, используя современные методы интеллектуального анализа данных.

Одно из проведенных исследований [2] результатов матчей Национальной хоккейной лиги (NHL) проводилось на данных, собранных за три месяца сезона 2012-2013, что включает в себя 517 игр. На основе этих данных было построено несколько статистических моделей: классическая, продвинутая и смешанная, которые отличались набором рассматриваемых хоккейных характеристик. К каждой модели были применены различные алгоритмы машинного обучения, для решения задачи классификации. Результаты исследования показали, что для поставленных задач, наиболее надежно себя показали нейронные сети.

При прогнозировании матчей Национальной футбольной лиги (NFL) исследователи применяли гибридный подход с использованием нейронной сети и метода линейной регрессии [3]. На первом этапе с использованием метода линейной регрессии определялся вес каждого статистического показателя, далее уже взвешенные показатели использовались нейронной сетью для предсказания победы. Утверждается, что такой подход дает высокую точность предсказания.

Проводимые исследования подтолкнули на проведение собственного анализа проблемы прогнозирования результатов матчей. В качестве объекта анализа был выбран статистический набор данных баскетбольных матчей Национальной баскетбольной ассоциации (NBA). В отличие от описанных ранее исследований, которые проводились на относительно небольшом временном отрезке, данные взяты на промежутке с 1985 по 2016 год, такой набор данных содержит информацию примерно о 32000 матчей.

Такой выбор обусловлен целью не только выявить показатели, в значительной степени определяющие успех команд, но и изучить тенденцию развития причин успеха со временем. Подобное исследование, например, поможет получить ответ на вопрос: «Какова вероятность того, что нынешние чемпионы, смогут обыграть лучшие команды двадцатилетней давности?».

В качестве инструментов проведения анализа были выбраны средства оперативной аналитической обработки данных (OLAP), однако в дальнейшем планируется также использование других методов анализа данных.

Основная идея OLAP заключается в построении многомерных кубов, которые будут доступны для пользовательских запросов. Исходные данные для построения таких кубов нередко содержатся в специализированных базах данных, называемых хранилищами данных.

Исследуемый набор данных содержит: дату проведения матча, место проведения, названия играющих команд, победителя, баскетбольную игровую статистику. Для формирования хранилища создается многомерная модель этих данных. Она состоит из таблицы фактов и таблиц измерений.

В таблице фактов содержатся сведения о каждом конкретном матче. В ее числовых полях приведены данные игровой статистики: она содержит в себе информацию о количестве набранных очков, проценте попадания, блокшотах, подборках, перехватах и других статистических параметров, содержащихся в исходном наборе данных.

Остальные данные отражены в описательных полях таблиц измерений. Из представленного набора данных можно выделить следующие измерения: время, команды, место проведения, игровые отрезки, в конкретном случае баскетбольный матч делится на 4 четверти. Итоговая структура хранилища представлена на рисунке 1.

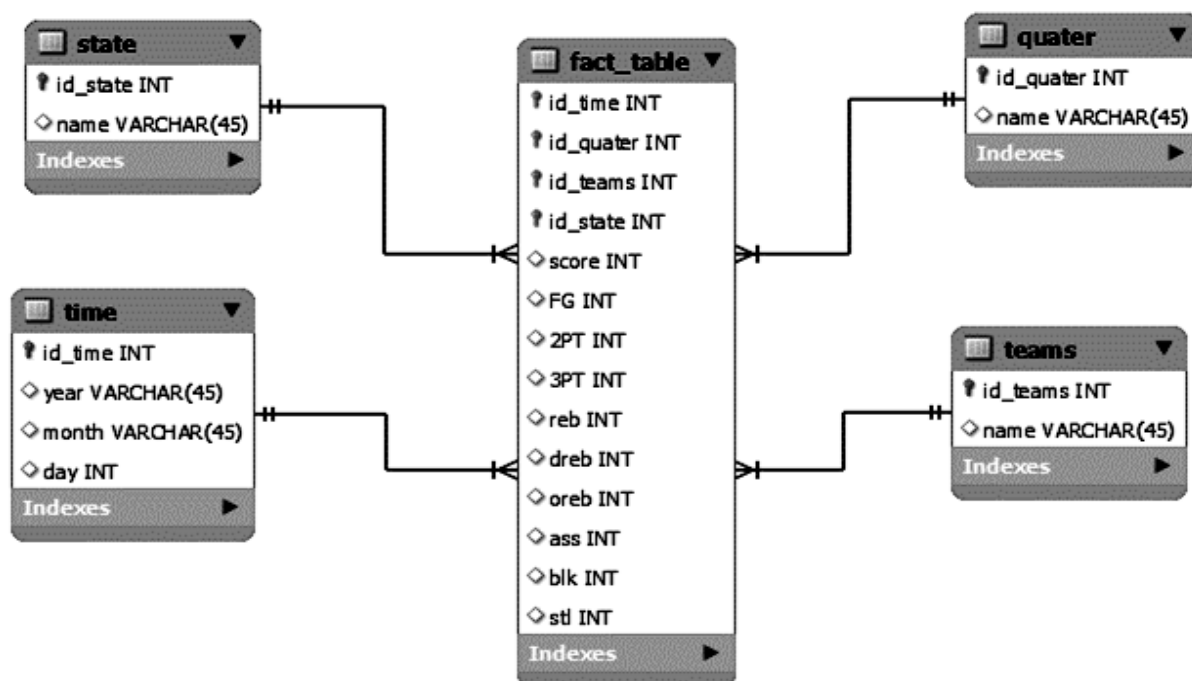


Рис. 1. Структура хранилища данных

На данном этапе работы, было завершено формирование хранилища данных, а также наполнение хранилища набором тестовых данных. Далее, при помощи плагина Saiku Analytics для OLAP сервера Pentaho, была реализована проверка гипотезы: «На домашней арене команды показывают результаты

лучше, чем в гостях». Saiku Analytics позволяет пользователю получить визуализированный результат проверки гипотезы, представленный на рисунке 2.

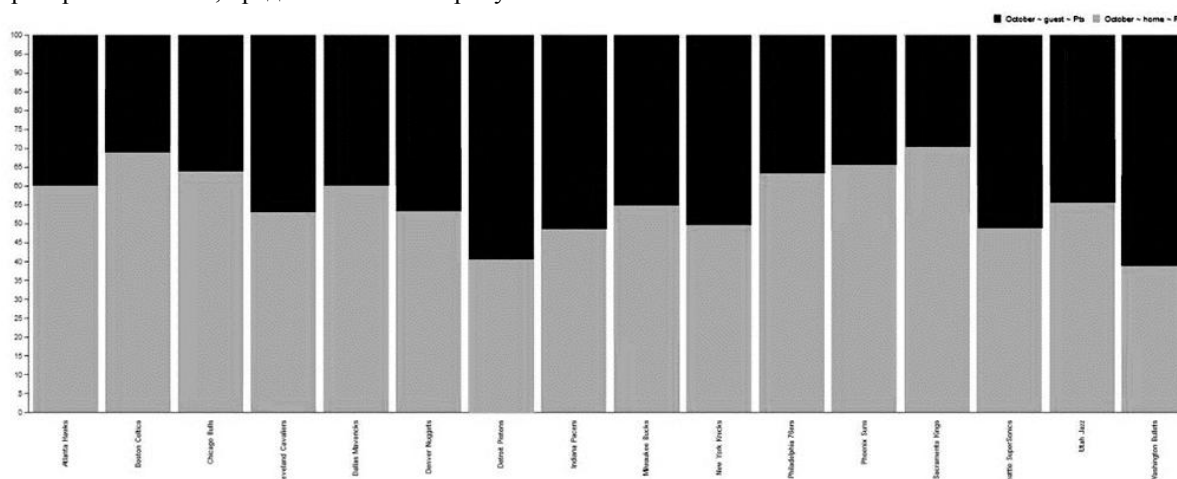


Рис. 2. Результат проверки гипотезы

Гипотеза оказалась справедливой для большинства команд лиги на рассматриваемом отрезке времени.

Дальнейшая работа будет направлена на усовершенствование структуры хранилища, заполнения его необходимым объемом данных, выполнение проверок новых гипотез, определение эффективного метода и проведение интеллектуального анализа данных.

В качестве результата ожидается получение определенных статистических характеристик, в большей степени влияющих на исход матча и использование этих характеристик, для формирования выводов о необходимых улучшениях в составах команд.

Список литературы

1. Бурева В.К., Стоянов Е.И. Применение методов искусственного интеллекта в спорте // Актуальные вопросы технических наук: материалы III Междунар. науч. конф. (г. Пермь, апрель 2015 г.). Пермь: Зебра, 2015. С. 1-12.
2. Weissbock J., Viktor H., Inkpen D. Use of Performance Metrics to Forecast Success in the National Hockey League. Machine Learning and Data Mining for Sports Analytics ECML/PKDD, 2013 workshop. Czech Republic.
3. Anyama O.U., Igiri C.P. An Application of Linear Regression & Artificial Neural Network Model in the NFL Result Prediction, International Journal of Engineering Research & Technology, 2015. № 4 (1). P. 457–461.