

АРХИТЕКТУРА СИСТЕМЫ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ ПО ОЦЕНКЕ РАЗВИТИЯ НАУЧНЫХ НАПРАВЛЕНИЙ

Солиев И.Б.

Томский политехнический университет, ОИТ ИШИТР, аспирант гр. А1-36,

e-mail: ibs2@tpu.ru

Аннотация

Предложена архитектура компьютерной системы поддержки принятия решений, позволяющая на основе собранных баз данных определить приоритетные научные направления, и выявлять новые тенденции.

Ключевые слова: система поддержки принятия решений, анализ данных, прогресс исследований, научные направления, стратегическое планирование.

Введение

В современном мире науки и технологий важность стратегического планирования исследований и разработок неоспорима. Эффективное распределение ресурсов, выбор приоритетных направлений для финансирования и поддержки зависят от точности и оперативности оценки развития научных направлений. Научно-исследовательские работы по проектированию СППР играют важную роль в этом улучшении, поскольку научные исследования по проектированию исследования могут вовлечь промышленность и специалистов в интеллектуально важные проекты. В данной работе актуальность определяется потребностью в систематизации больших объемов научных данных и выработке на их основе обоснованных прогнозов. Целью работы является создание архитектуры системы поддержки принятия решений, которая позволит ученым, управленцам и исследователям принимать взвешенные решения о развитии научных направлений. Новизна заключается в применении современных алгоритмов обработки данных и машинного обучения для повышения точности прогнозов.

Описание алгоритма

Архитектура системы поддержки принятия решений (СППР) по оценке развития научных направлений представляет собой сложную информационную и аналитическую структуру, предназначенную для обработки больших объемов данных, их анализа и предоставления рекомендаций для различных заинтересованных сторон, включая научные учреждения, правительственные органы, частных инвесторов и исследователей [1].

Таблица 1

Компоненты архитектуры СППР

ФУНДАМЕНТАЛЬНЫЕ КОМПОНЕНТЫ АРХИТЕКТУРЫ СППР	
Модуль сбора и предварительной обработки данных	Автоматический сбор данных из различных источников, таких как научные публикации, патенты, научные журналы, базы данных исследований и разработок, а также социальные сети и новостные потоки, где обсуждаются научные темы. Эти данные могут включать текст, изображения, аудио и видеоматериалы, а также структурированные данные, например, результаты экспериментов или статистику финансирования. Предварительная обработка включает очистку данных, их нормализацию, а также их классификацию и аннотирование с использованием методов машинного обучения и естественной обработки языка [2].
Модуль аналитики и обработки данных	Модуль включает в себя алгоритмы машинного обучения, статистический анализ и моделирование, которые применяются к собранным данным для выявления тенденций, закономерностей и взаимосвязей. Он может использовать различные подходы, включая предиктивную аналитику, кластерный анализ, сетевой анализ и анализ временных рядов для оценки текущего состояния и потенциального развития научных направлений [2].
Модуль поддержки принятия решений	Основывается на результатах предыдущих модулей и включает инструменты для создания сценариев развития, оценки рисков и выработки оптимальных стратегий. Модуль может включать экспертные системы, системы, основанные на правилах, а также интеграцию с инструментами искусственного интеллекта для генерации рекомендаций [2].

Алгоритм функционирования СППР

АЛГОРИТМ ФУНКЦИОНИРОВАНИЯ СППР	
Инициализация системы	Запуск системы начинается с конфигурации параметров сбора данных и определения целей анализа. Это включает настройку фильтров данных, определение ключевых слов, научных областей и задач, которые система должна решать [3].
Сбор данных	Система автоматически собирает данные из предварительно определенных источников, используя средства веб-скрапинга, API доступа к базам данных и другие методы сбора информации [3].
Предварительная обработка данных	Собранные данные очищаются от шума, дубликатов и несущественной информации. Производится их структурирование и аннотация с помощью алгоритмов машинного обучения для дальнейшего анализа [3].
Анализ данных	С помощью статистических методов, алгоритмов машинного обучения и моделей, данные анализируются на предмет выявления закономерностей и трендов в развитии научных направлений. Анализ может включать оценку цитируемости, коллабораций между исследователями, распределение финансирования и другие параметры [3].
Прогнозирование и моделирование	На основе исторических данных и текущих тенденций система строит прогнозы развития научных направлений. Моделирование различных сценариев помогает оценить потенциальные риски и возможности [3].
Визуализация данных	Результаты анализа и прогнозов представляются в виде интерактивных диаграмм, карт и отчетов для удобства интерпретации и принятия решений [3].
Поддержка принятия решений	Система предоставляет рекомендации и стратегические советы на основе сценарного анализа, оценки рисков и оптимизации. Это может включать предложения по фокусировке на определенных областях исследований, инвестициях в научные проекты или разработке новых политик [3].
Итерация и обучение	Система постоянно обновляется с новыми данными, что позволяет улучшать алгоритмы и уточнять модели. Обратная связь от пользователей и экспертов используется для корректировки и улучшения процессов СППР [3].

Таблица 3

Технологические аспекты реализации СППР

ТЕХНОЛОГИЧЕСКИЕ АСПЕКТЫ РЕАЛИЗАЦИИ СППР	
Cloud services	Использование облачных сервисов для масштабирования ресурсов и обеспечения высокой доступности системы [4].
Big Data	Применение технологий больших данных (Big Data) для обработки и хранения большого объема разнообразной информации [4].
UI/UX Interface	Разработка пользовательского интерфейса (UI/UX), который обеспечивает легкость использования и понимания системы пользователями разного уровня [4].

Архитектура системы поддержки принятия решений представлена на рис. 1.

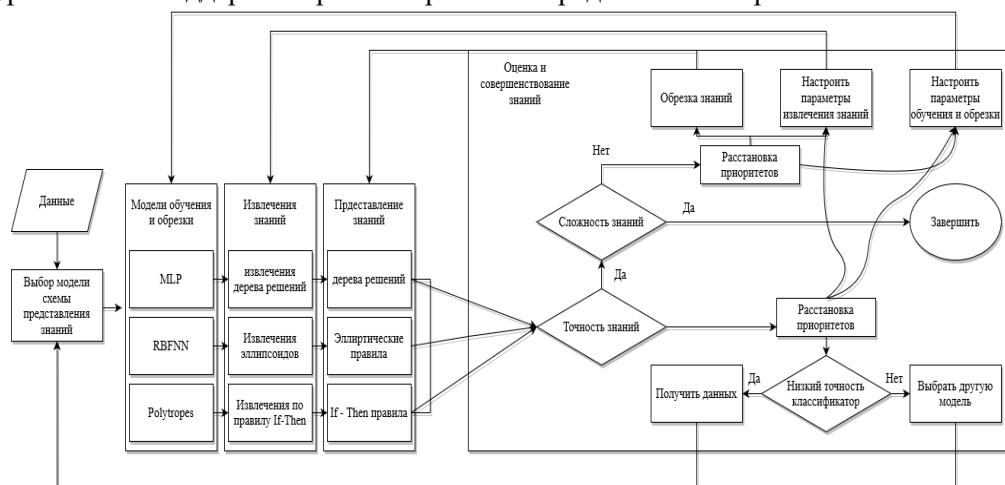


Рис. 1. Архитектура системы поддержки принятия решений

При первоначальном выборе модели классификатора следует учитывать размерность классифицируемого набора данных. Наборы данных малой размерности могут обрабатываться RBFNN (радиально базисная нейронная сеть) [5]. Для случаев с десятью или меньшими входными размерностями, можно построить линейный выпуклый классификатор многогранников (или основанный на много векторном методе древовидный классификатор) и применить разработанный метод извлечения правил, основанный на оптимизации. Наконец, двоичное дерево решений по классификации может быть извлечено из ИНС, независимо от числа входных размерностей. Кроме того, нет никаких теоретических ограничений для применения алгоритма извлечения дерева решений к RBFNN. Но, в любом случае, согласно теореме, не может быть единой классификационной модели, одинаково хорошо работающей на всех видах наборов данных. Таким образом, прямой ИНС не может быть лучшим классификатором во всех случаях; поэтому могут использоваться классификаторы RBFNN и многогранники.

Модели подготовки и обучения. Подготовка моделей выходит за рамки настоящего документа, однако в рамках извлечения дерева решений из многослойного полностью подключенного ИНС было разработано сокращение модели. Тем не менее, сети RBFNN (в эллиптическом пути извлечения правил) можно сократить с помощью разработанного алгоритма [6]. Сокращение модели является первым шагом, на котором можно контролировать сложность извлеченных знаний. Будет использоваться более агрессивная обрезка, меньшее дерево решений будет извлечено. Менее сложные правила / знания означают более понятную и менее точную – в зависимости от сценария использования знаний, агрессивное сокращение может быть желательным или нет.

После выполнения модели обучения можно провести этап извлечения знаний. Многослойные ИНС могут быть представлены как двоичная классификация дерева решений, классификатор RBFNN можно перевести в эллиптический классификатор (Кроме того, нет никаких теоретических ограничений для применения метода извлечения дерева решений, но никаких экспериментов еще не проводилось). В конце концов, If – Then правила можно извлечь из фрагментарного линейного классификатора [7].

Оценка и совершенствование знаний. Заключительный этап, который следует выполнить, – оценка полученного знания и улучшение потенциальных проблем. На рис. 2 используется такая формулировка, как «точность слишком низка», «знание слишком сложно» - «мы должны остановиться здесь и описать корни такого неофициального описания. Как обычно, конечные цели стимулируют процесс, извлечения знаний, если требуется понимание модели (например, как именно конкретный классификатор принимает решение об отказе в предоставлении кредита) можно предположить, что меньшее число правил имеет большее значение; таким образом, падение точности модели приемлемо [8]. Допустимая величина падения точности также относится к конкретному случаю. В одном случае допустимо снижение точности классификации на 10 %, а в других - снижение на 2 %. То же самое относится и к сложности знаний - является ли она сложной или не зависит от субъективного суждения экспертов по доменам. Первым шагом в оценке и уточнении является проверка эффективности классификации. Если эффективность классификации является слишком низкой, существует несколько вариантов: во-первых, параметры извлечения знаний могут корректироваться; это применимо к процедуре извлечения правила Если-тогда, поскольку она имеет порог глубины рекурсии и применяется к процедуре извлечения эллиптического правила, так как она может иметь порог для подсчета извлеченных эллипсоидов. Оба порога могут быть увеличены, чтобы получить больше грубых правил. Если это не так, то параметры обрезки можно ослабить, а обрезку повторить (с последующим извлечением знаний).

Если процедура обрезки является чрезмерной, точность модели снижается слишком сильно. Наконец, если предыдущие шаги не помогают, то саму модель следует переучивать, но только если мы увидим, что разница между точностью классификатора и точностью знания невелика. Таким образом, получив более эффективный классификатор, мы можем получить более эффективные знания (с точки зрения классификации). В случаях, когда эффективность знаний является хорошей, но знания не понятны, то есть слишком сложны - существует несколько вариантов. Во-первых, знания обрезка может быть выполнена. В случае дерева решений по классификации могут применяться стандартные существующие подходы к сокращению дерева решений. В случае If-Then или эллиптических правил, ручной выбор наименее важных правил должен выполняться с последующим устранением. Наиболее очевидными критериями для наименее важного правила могут быть либо подсчет охватываемых учебных проб данных, размер района, охватываемого правилом, либо сочетание обоих критериев [9].

Если такое сокращение знаний не удастся, то либо параметры извлечения знаний следует изменить на более строгие, или обрезка сама должна быть настроена для удаления большего количества нейронов (или получение меньшего числа многогранников). Наконец, если ничего не помогает, то саму модель следует переобучить; конечно, ее следует упростить, с тем чтобы мы извлекали менее сложные знания.

Альтернативный курс действий должен быть выполнен в случае, если предыдущие шаги не принесли результатов. Альтернативой является оценка эффективности модели в отношении входных данных если имеется только конкретный небольшой точность входного пространства данных, который плохо классифицирован, то только для этого конкретного региона можно обучить другой классификатор и извлечь знания для этого подмножества данных. Если это не так, то весь классификатор следует отбросить и ввести новый цикл модельной подготовки, сокращение, извлечение знаний и проведение оценки.

Заключение

Архитектура системы поддержки принятия решений, описанная в данной работе, представляет собой важный инструмент для современной науки. Ее эффективное функционирование обеспечивает точность аналитических данных, способствующих принятию обоснованных решений в области науки и технологий. Развитие и дальнейшее совершенствование подобных систем станет ключевым фактором в успешном развитии научных направлений и технологических инноваций в будущем.

Список использованных источников

1. Alavi M., & Joachimsthaler E.A. Revisiting DSS implementation research: A meta-analysis of the literature and suggestions for researchers // *MIS Quarterly*. – 1992. – 16(1). – P. 95–116.
2. Adkins M., Burgoon M. and Nunamaker J.F. Using Group Support Systems for Strategic Planning with the United States Air Force // *Decis Support Syst.* – 2003. – Vol. 34(3).
3. Burstein F., & Carlsson S.A. Decision Support through Knowledge Management. In F. Burstein & C.W. Holsapple (Eds.), *Decision support systems handbook*. Berlin: Springer-Verlag. – 2008. – P. 103–120
4. Eierman M.A., Niederman F., & Adams C. DSS theory: A model of constructs and relationships // *Decision Support Systems*. – 1995. – 14(1). – P.1-26.
5. Braun R.L., & Davis H.E. Computer-assisted audit tools and techniques: Analysis and perspectives // *Managerial Auditing Journal*. – 2003.
6. Alles M., & Gray G. The use of CAATTs in auditing financial statements: Implications for auditing practice and education // *Journal of Accounting Education*. – 2016. – 36. – P.1–17.
7. Manita R., Berrada A., & El Amrani El Idrissi, N. The impact of information technology on audit quality: A literature review // *Journal of Accounting and Taxation*. – 2020. – Vol. 12(1). – P. 1–13.
8. Tiberius V., & Hirth M. The role of information technology in enabling open innovation: Complementing absorptive capacity // *Journal of Business Research*. – 2019. – Vol. 94, – P. 203–215.
9. Andrey Bondarenko, Methodology for Knowledge Extraction from Trained Artificial Neural Networks // *Information Technology and Management Science*. – 2018. – Vol. 21. – P. 6–14.