

МЕТОДИКА «ИНТЕЛЛЕКТУАЛЬНОЙ» ОБРАБОТКИ СОЦИАЛЬНО-ЭКОНОМИЧЕСКОЙ ИНФОРМАЦИИ В АВТОМАТИЗИРОВАННОЙ СИСТЕМЕ ПОДДЕРЖКИ НАУЧНЫХ ИССЛЕДОВАНИЙ

Измайлов В. Г., Озерова М. Ю., Трейбач А. Л.

В процессе проектирования и разработки отдельных модулей автоматизированной системы поддержки научных исследований в области экономики образования (АСПНИ) предложена и реализована методика «интеллектуальной» обработки статистической информации.

Возможность использования известных методов математической статистики и машинного обучения для решения задач анализа сверхбольших объемов информации открыла новые возможности перед аналитиками, исследователями, а также теми, кто принимает решения — менеджерами и руководителями компаний. Сложность и разнообразие методов интеллектуального анализа требуют создания специализированных средств конечного пользователя для решения типовых задач анализа информации в области бизнеса, экономики, науки.

Перед исследователями часто возникает проблема обработки огромного количества информации, так или иначе касающейся темы исследования. Научная студенческая группа факультета информатики Магнитогорского государственного университета (МаГУ) D-Studio проводит исследовательские работы в области экономики образования. Общая цель научного исследования: изучение процессов в современной экономике знаний и построение моделей, адекватно отражающих особенности этих процессов для выработки рекомендаций по принятию решений.

Для достижения этой цели в D-Studio ведутся также и экспериментальные работы по «интеллектуальной» обработке статистической информации. Для анализа статистических данных D-Studio разрабатывает автоматизированную систему поддержки научных исследований (АСПНИ).

Данная разработка представляется актуальной в связи с наличием следующих факторов:

- современные инструментальные средства поддержки принятия решений должны «уметь» обрабатывать «нечеткую», «неточную», слабоструктурируемую информацию, т.к. именно такая информация служит в большинстве случаев основой для принятия решений;
- в различных экономических процессах, в частности, в образовательных процессах, информационная составляющая на современном этапе играет все более значительную роль и может как «улучшить» так и исказить отражение процесса в восприятии активных агентов (например, лиц, принимающих решение);
- разрабатываемый модуль в архитектуре АСПНИ сможет, с одной стороны, обеспечить инструментарий для дальнейшей обработки «нечеткой» информации, и с другой стороны, повысить качество представления исследователя об изучаемой проблемной ситуации.

Проведение любого анализа требует прохождения определенных шагов, по окончании которых исследователь достигает желаемого результата, то есть может сделать вывод о сложившейся ситуации.

Студенческая научная группа D-Studio при анализе проблемной ситуации в области экономики образования предложила методику (**методика** — это набор шаблонов, процедур и рецептов, конкретизирующих метод) «интеллектуальной» обработки социально-экономической информации, которую реализует в автоматизированной системе поддержки научных исследований, тем самым, проверяя ее на практике.

Методика «интеллектуальной» обработки социально-экономической информации подразумевает решение следующих методических задач.

Во-первых, при анализе социально-экономической информации, как и в других ситуациях, необходимо поставить цель исследования, потому что именно правильная постановка цели определяет то, какие результаты необходимо получить по окончании исследования. Наличие точных, определенных, ясно поставленных целей позволяет добиться наиболее эффективных результатов.

Цель работы D-Studio: разработка модуля автоматизированной системы поддержки научных исследований (АСПНИ), выполняющего функции поиска в распределенных базах данных (например, в internet) информации, отвечающей заданным критериям («нечетким»), импорт

данных в БД АСПНИ и статистический анализ найденной социально-экономической информации с целью принятия решений в исследуемой предметной области (Лисьев, 2008).

Для достижения поставленной цели необходимо провести оценку форматов, качества исходной статистической информации. В процессе исследования было установлено, что найденная статистическая информация неоднородна и имеет различную структуру и наполнение, а также встречаются «пробелы» и некорректные значения в данных. Возникла необходимость создания специального приложения по работе с полученной «сырой» и не готовой к обработке информацией. Для решения данной проблемы было разработано приложение по удалению неверных (некорректных) значений, которые не могут участвовать в дальнейшем анализе данных (символьные значения, отрицательные значения с неверным форматом); удалению кортежей с пустыми и/или некорректными значениями; сортировке данных и т.п.

Следующим этапом является разработка модельных онтологий для использования в качестве тестовых «баз знаний» для поиска и импорта информации. Для решения задачи «интеллектуальной» обработки данных перспективным является применение мультиагентного подхода, представляется актуальным создание информационной системы поддержки принятия решений на основе многоагентного подхода.

Технология многоагентных систем — это новая парадигма информационной технологии, ориентированной на совместное использование научных и технических достижений и преимуществ, которые дают идеи и методы искусственного интеллекта, современные локальные и глобальные компьютерные сети, распределенные базы данных и распределенные вычисления, аппаратные и программные средства поддержки теории распределенности и открытости. Принципиальным отличием новой парадигмы построения прикладных систем является то, что в ней определяющим являются данные, которые указывают направление вычислений (Андреев с соавт., 2007).

В процессе разработки модуля интеллектуальной обработки данных в системе поддержки научных исследований перед D-Studio возникла проблема извлечения и централизации разнородных знаний, необходимых для организации работы интеллектуальных агентов, и их эффективного использования. Актуальными стали вопросы о том, каким образом можно собрать и структурировать знания одной предметной области, а также как организовать доступ агентов к знаниям. Для реше-

ния данной проблемы было решено использовать технологию онтологического анализа.

Онтология определяет термины, которые позволяют описать предметную область. Использование онтологий необходимо в приложениях-агентах, осуществляющих поиск и объединение информации из различных источников и из разных сред, в которых один и тот же термин может означать разные вещи. Разработка онтологии сродни определению набора данных и их структуры для использования другими программами. Методы решения задач, доменно-независимые приложения и программные агенты используют в качестве данных онтологии и базы знаний, построенные на основе этих онтологий. На данном этапе в D-Studio разрабатывается онтология предметной области «экономика образования». Данная онтология будет являться базой знаний для организации функционирования интеллектуальных агентов, целью которых будет сбор и обработка необходимой исследователю информации (Абрамов с соавт., 2008).

После создания онтологической базы знаний, на основании которой возможно принимать решение, необходимо спроектировать и разработать модуль, выполняющего функции поиска в распределенных базах данных информации, отвечающей заданным критериям. Разработка данного модуля позволит «интеллектуализировать» процесс поиска необходимой для исследования информации, что в свою очередь освободит исследователя от необходимости осуществлять данный поиск вручную и тем самым позволит сократить временные затраты. Для разработки модуля «интеллектуального» поиска D-Studio планирует использовать технологию мультиагентных систем.

Технология их создания основывается как на базе новых аппаратных решений, так и новой концепции управления и взаимодействия между программными компонентами, получившими название интеллектуальных агентов. Каждый такой агент обеспечивает управление доступными для него информационными ресурсами, используя для этого набор собственных целевых условий и внешних задающих воздействий. Сами агенты, как программные компоненты единой сложной системы, передают данные и исполняемый код алгоритмов управления по сети с помощью стандартных средств межсетевое взаимодействия. Таким образом, в механизм управления добавляется новая возможность воздействия на состояние объекта путем передачи не самого регулирующего сигнала, а программной компоненты, реализующей алгоритм управления. В результате эта компонента получает оперативные данные о со-

стоянии объекта уже непосредственно в точке, где необходимо реализовать требуемое воздействие. При этом в алгоритме могут учитываться координирующие воздействия от других компонент.

Управление на базе такой технологии носит ярко выраженный иерархический характер, при котором на нижнем уровне реализуется механизм управления с обратной связью и малыми задержками, а на верхнем — параметрическая оптимизация и программная координация управляющих воздействий.

Разработанная таким образом система, позволит осуществлять быстрый и гибкий поиск необходимой информации, в том числе и статистической, отвечающей заданным критериям. Найденная информация обрабатывается, оцениваются форматы и ее качество.

На следующем этапе «чистая» информация экспортируется в базу данных АСПНИ для хранения и последующей статистической обработки и анализа. Создание БД позволяет осуществлять выборку необходимой исследователю информации из массива найденной «интеллектуальным» агентом.

Необходимым условием получения результата при использовании данной методики является реализация приложения в выбранной программной среде. D-Studio разрабатывает автоматизированную систему поддержки научных исследований в программной среде PHP и на СУБД MySQL. Для разработки онтологий и мультиагентных систем выбрана инструментальная среда MagentA Engine RC компании Маджента, Россия. Данный «метаинструмент», обеспечивает возможность построения прикладных мультиагентных систем с жестко закодированной онтологией предметной области и помогает пользователям создавать системы приобретения знаний для конкретной предметной области.

По итогам проведенных работ, статистическим результатам производится тестирование и оценка эффективности предложенных решений. Оценка производится экспертами, для того чтобы оценка оказалась верной и адекватной, необходимо наличие у экспертов определенных знаний и навыков. В D-Studio экспертами являются студенты-участники научной группы. Существуют различные методики подготовки специалистов в области информационных технологий (обмен опытом, приглашенные специалисты, создание научных центров и др.), но большинство из них являются масштабными, требующими привлечения большого числа специалистов и материальных ресурсов. В D-Studio практикуется метод проектного обучения — метод предполагающий организацию

деятельности, направленную на получение результата (продукта), в ходе которого идет усвоение новых знаний и действий, он интересен и эффективен подлинностью получаемого опыта. Студент получает не только теоретические знания, но и имеет возможность применить их на практике. В процессе разработки происходит ликвидация пробелов в знаниях, а также их углубление. Проектный метод позволяет отойти от авторитарности в обучении, всегда ориентирован на самостоятельную работу в сочетании с групповой организацией деятельности, что позволяет приобретать коммуникативные навыки и умения (умение работать в группах в разнообразных качествах, использовать широкие человеческие контакты, разные точки зрения на одну проблему). Это становится тем актуальней, чем интенсивней развиваются средства информационных коммуникаций, деловое и культурное сотрудничество между людьми — носителями разных культур мирового сообщества.

Метод проектов является активным способом обучения. В процессе работы над проектом происходит тесное личностное взаимодействие ученика с учителем на принципах равного партнерства, общение старшего по опыту товарища с младшим с одновременным отсутствием диктата со стороны учителя и достаточной степенью самостоятельности для ученика. Целью организованной деятельности является получение интересного для обучаемого результата — результата работы над проектом — что является самым сильным мотивантом.

Большинство научных исследований подразумевает групповую работу и требует наличия способов передачи данных между участниками проектных групп, что способствует получению результатов. Для организации групповой работы участниками D-Studio разработан специальный программный модуль, позволяющий создавать проектные группы. В рамках проекта можно определять его участников, разграничивать права доступа к данным. Каждый участник может загружать файлы проекта для организации коллективной работы и обсуждения результатов, тем самым, облегчая и ускоряя процесс работы.

Следование методике позволило участникам D-Studio спроектировать автоматизированную систему поддержки научных исследований, которая позволит значительно упростить дальнейшую исследовательскую работу и позволит провести анализ проблем в области экономики образования, выработать рекомендации для их разрешения.

СПИСОК ЛИТЕРАТУРЫ

- Абрамов Д.В., Андреев В.В., Симонова Е.В., Скобелев П.О.* Открытые мультиагентные системы для принятия решений в задачах динамического распределения ресурсов. — Самара: Поволжская государственная академия телекоммуникаций и информатики, 2008.
- Андреев В.В., Минаков И.А., Пшеничников В.В., Симонова Е.В., Скобелев П.О.* Основы построения мультиагентных систем. — Самара: Поволжская государственная академия телекоммуникаций и информатики, 2007.
- Лисьев Г.А.* Автоматизированная система поддержки научных исследований в области экономики образования II Труды Международных научно-технических конференций «Интеллектуальные системы» (AI'08) и САПР» (САИТ-2008) // Научное издание в 4-х томах. — М.: Физматлит, 2008. — Т. 1. — С. 280–285.

METHOD FOR «INTELLIGENT» PROCESSING OF SOCIO-ECONOMIC INFORMATION IN SCIENTIFIC RESEARCH SUPPORT SYSTEM

Izmaylov V., Ozerova M., Treybach A.

Technique for “intelligent” processing of statistical information is offered and implemented as a part of the scientific research support system (SRSS). The system was probed on economical data of educational sector.