

Системное представление технологии интеллектуального анализа данных

И. А. Брусакова

Зав. кафедрой инноватики
Санкт-Петербургский государственный
электротехнический университет
«ЛЭТИ» им. В.И. Ульянова (Ленина)
brusakovai@mail.ru

В. Д. Чертовской

Профессор кафедры информационных систем
Санкт-Петербургский государственный
электротехнический университет
«ЛЭТИ» им. В.И. Ульянова (Ленина)
vdchertows@mail.ru

Аннотация. Отмечено все более широкое распространение технологии интеллектуального анализа данных (ИАД). Показана потребность в упорядочении большого количества опубликованных материалов. Предложен и выполнен вариант систематизации с помощью системного анализа, что упрощает процедуры создания (синтеза) и исследования систем на основе технологии ИАД. Рассмотрен один из алгоритмов синтеза.

Ключевые слова: технология; системное представление; интеллектуальный анализ данных; упорядочение

I. ВВЕДЕНИЕ

Технология интеллектуального анализа данных получает все более широкое распространение. В настоящее время по интеллектуальному анализу данных имеется значительное количество публикаций, освещающих различные аспекты предметной области. Существует множество методов реализации технологии, что затрудняет понимание и усвоение материала, анализ и синтез систем с применением этой технологии.

Необходима систематизация, системное рассмотрение материалов, накопленных в этой только что начавшей складываться предметной области.

Именно эта задача решается в настоящей работе. Преследуется цель системного формирования теоретических и прикладных знаний для последующего создания (синтеза) и исследования сложных систем управления организационными объектами. В силу начального этапа развития рассматриваемого класса технологий значительное место отведено неформальным методам описания процессов.

II. ПОСТАНОВКА ЗАДАЧИ

Известно, что интеллектуальный анализ данных (ИАД) – технология исследования данных, использующая методы искусственного интеллекта и ориентированная на придание системе свойств искусственного интеллекта. Эту технологию называют и Data Mining, которая представляет фактически использование существующих методов интеллектуальных систем.

Для систематизации многочисленных формальных методов [1–5], присущих технологии ИАД, используем метод системного анализа.

III. РЕШЕНИЕ ЗАДАЧИ

Предложено системное упорядочение методов по схеме «интеллектуальные системы – локальные технологии – процедуры ИАД». Эта схема (рис. 1) определяет этапы технологии создания систем с ИАД.

На первом этапе формируется цель создания системы, на основе которой проводится выбор одной из интеллектуальных систем: многоагентные системы [2], интеллектуальные системы управления, экспертные системы, искусственные нейронные сети [1], системы с генетическими алгоритмами [4], системы на естественном языке. Выбор определяется возможностями систем. Приведем их назначение.

Многоагентная (мультиагентная) система – совокупность взаимосвязанных агентов как программных, так и аппаратных средств, способных взаимодействовать друг с другом и окружающей средой, обладающих определенными интеллектуальными способностями и возможностью индивидуальных и совместных действий. Интеллектуальная система управления – система, работающая в режиме целевой адаптации. Экспертная система – вычислительная система, использующая знания эксперта, процедуры логического вывода и позволяющая дать объяснения полученным результатам. Системы на естественном языке – поисковые и алгоритмические системы, работающие с текстовой информацией. Искусственные нейронные сети – системы, имитирующие свойства нейрона.

На втором шаге (рис. 2) производится определение применяемой локальной технологии [1]: реальное время, анализ текста; извлечение знаний из Web; визуальный анализ.

Технологии реального времени является дополнением к совокупности статистических методов. Они используют технологии накапливаемого обучения (к анализируемым данным постоянно добавляются новые) при

прогнозировании возможно с применением обратной связи.

Технология анализа текста позволяет выявить новые, полезные шаблоны в неструктурированных данных, получаемых из баз данных. При этом требуется предварительная обработка текста.

Технология извлечение знаний из Web, являясь фактически разновидностью технологии анализа специфической текстовой информации. При большом количестве информации в Интернете обычный поиск связан с серьезными трудностями. К ним относятся: малый процент нужной информации; редкая повторяемость вызовов; персонализация и настройка системы на конкретного пользователя. Технология получения информации в этом случае включает поиск ресурсов, извлечение информации по ключевым словам, обобщение и анализ результатов. Извлечение информации проводится с участием человека.

Визуальный анализ проводится при создании практически любой системы с использованием ИАД. Цель визуального анализа – представление данных в форме, более наглядной и удобной для работы пользователя. Дело в том, что при большом количестве данных новые знания не всегда можно найти простейшими средствами графики (деревья решений, дейтограммы, двумерные графики) и требуются более сложные средства отображения. Необходим диалоговый режим работы пользователя, который дает возможность работать с неоднородными и сильно зашумленными данными, не применяя очень сложные алгоритмы решения. При этом повышается скорость обработки данных и конфиденциальность результатов.

На третьем этапе выбираются конкретные процедуры ИАД: классификация, кластеризация, выделение последовательностей и ассоциаций, прогноз. Каждая из процедур может быть выполнена различными математическими методами. Их классификация может быть такой (рис. 1).

Прогноз

Функции матричного аргумента
Многомерный нормальный закон
Оптимальный среднеквадратический линейный прогноз
Моделирование многомерных случайных данных

Визуализация

Главные компоненты и факторный анализ
Дискриминантный анализ
Нелинейные главные компоненты
Визуальный анализ данных – Visual Mining

Классификация и кластер

Стохастические расстояния и меры близости
Кластерный анализ
Метод Naïve Bayes
Деревья решений
Машины опорных векторов (Support Vector Machines)
Адаптивные методы кластеризации

Последовательности

Непараметрические методы
Параметрические модели
Прогнозирование
Статистическое исследование зависимостей
Метод наименьших квадратов
Ортогональная регрессия
Множественная регрессия
Нелинейные методы
Обработка данных с пропусками

Ассоциации

Поиск ассоциативных правил

Оцифровка неколичественных данных

Экспертные методы
Метод анализа иерархий Саати
Метрическое и неметрическое шкалирование

Анализ текстовой информации

Извлечение ключевых понятий
Классификация и кластер
Аннотирование
Особый класс составляют специфические методы реализации ИАД.

Развитие концепции баз данных

Системы поддержки принятия решений
Хранилища данных
OLAP-системы
Стандарты Data Mining

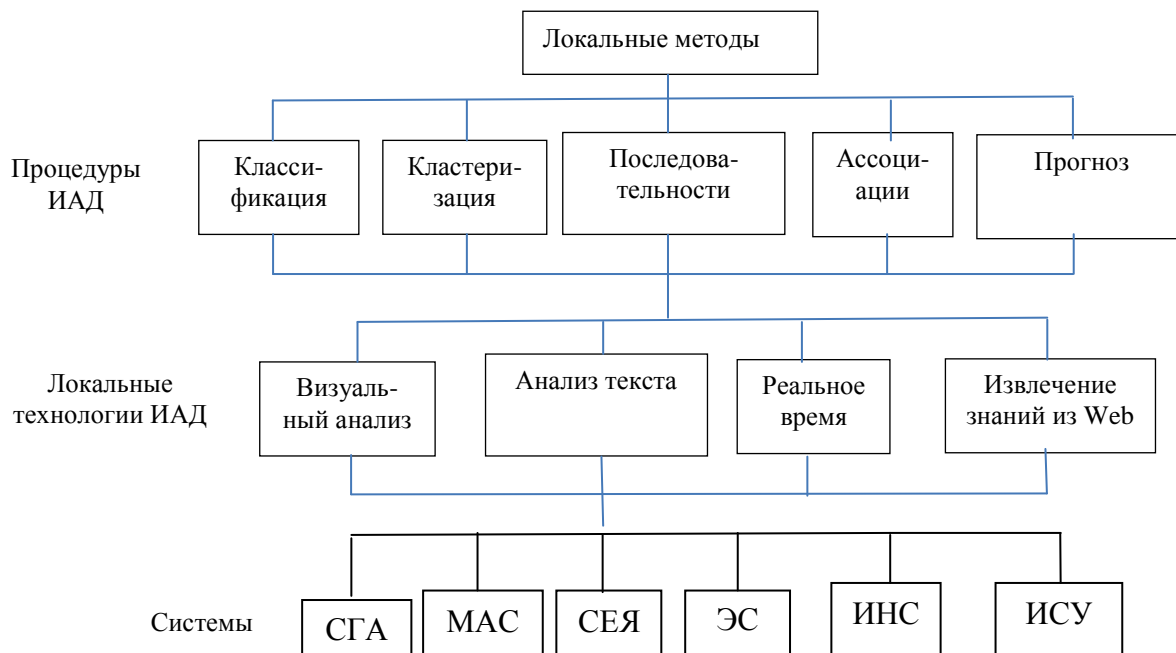


Рис. 1. Система составных частей технологии ИАД: СППР – системы поддержки принятия решений; СГА – системы с генетическими алгоритмами; МАС – многоагентные системы; СЕЯ – системы на естественном языке; ЭС – экспертные системы; ИНС – искусственные нейронные сети; ИСУ – интеллектуальные системы управления

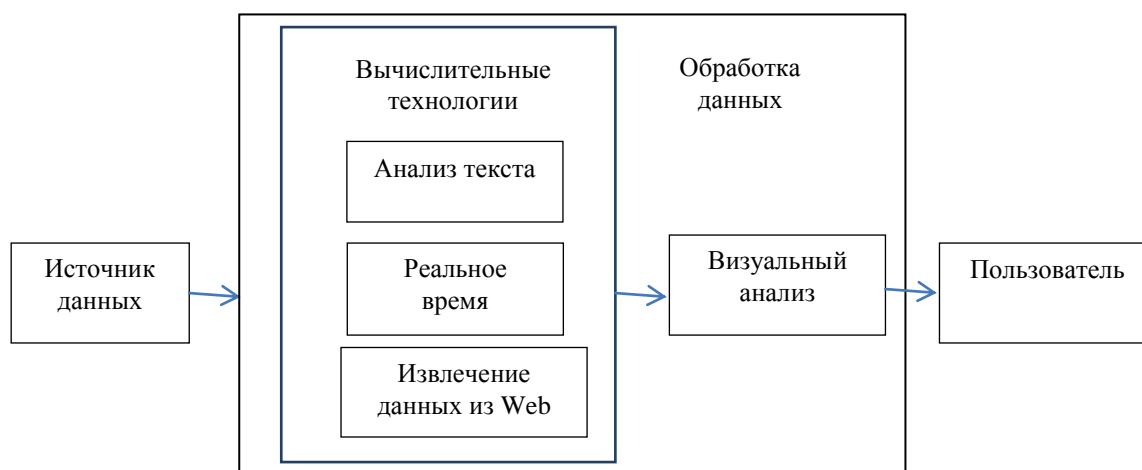


Рис. 2. Схема использования локальных технологий ИАД

IV. ПРИМЕР ПРИМЕНЕНИЯ ТЕХНОЛОГИИ ИАД

Чаще всего технологию ИАД иллюстрируют с применением многоагентных систем [1] Далее рассмотрим использование для названных целей интеллектуальных систем управления [2].

Рассмотрим автоматизированную систему управления производством, состоящую из двух подсистем: технико-экономического планирования и оперативного управления основным производством. Система имеет трехуровневую структуру управляющей части: «руководитель – диспетчер – начальники цехов». Для математического описания такой системы предложен и апробирован [2] однородный метод, построенный на базе задачи динамического

линейного программирования (ДЛП). Задача ДЛП составлена из широко известной задачи статического линейного программирования и системы разностных уравнений. На основе математического описания строится математическая модель, работающая в оптимальном режиме и позволяющая проводить процесс планирования как в традиционном варианте, так и при оперативном переходе на выпуск новой продукции. Последнее свойство позволяет изучать процедуру прогнозирования на продолжительных интервалах времени по схеме «что будет, если?», т.е. использовать технологию ИАД на основе интеллектуальной системы управления. В качестве локальной технологии выбрана технология реального времени, а в качестве процедур – прогнозирование.

V. ЗАКЛЮЧЕНИЕ

В настоящее время наиболее развитыми классами систем являются системы ИАД на базе многоагентных систем, интеллектуальных систем управления, отчасти – экспертных систем. Вместе с тем представляет интерес использование технологии в искусственных нейронных сетях, системах с генетическими алгоритмами, системах на естественном языке.

Дальнейшее развитие теоретических аспектов интеллектуального анализа данных связано как с совершенствованием пропозициональной логики, так и логик, более близких к психологии выработки знаний человеком.

СПИСОК ЛИТЕРАТУРЫ

- [1] Барсегян А.А. Анализ данных и процессов: учеб. пособие / А.А. Барсегян, М.С. Куприянов, И.И. Холод, М.Д. Тесс, С.И. Елизаров. СПб.: БХВ, 2009. 512 с.
- [2] Советов Б.Я., Цехановский В.В., Чертовской В.Д. Интеллектуальные системы и технологии. М.: Академия, 2013. 320 с.
- [3] Шабанов В.И., Андреев А. М. Метод классификации текстовых документов, основанный на полнотекстовом поиске // Труды первого российского семинара по оценке методов информационного поиска. Под ред. И.С. Некрестьянова. СПб.: НИИ Химии СПбГУ, 2003.
- [4] Мхитарян В.С. Анализ данных. М.: Юрайт, 2016.
- [5] Олифер Н., Олифер А. Сетевые операционные системы. СПб.: Питер, 2001. 525 с.