

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ПРОЕКТИРОВАНИИ И ПРОИЗВОДСТВЕ СЛОЖНЫХ СИСТЕМ

УДК 001.891.5

АЛГОРИТМЫ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ В ИНФОРМАЦИОННОЙ СИСТЕМЕ ПОДДЕРЖКИ УДАЛЕННОГО ЭКСПЕРИМЕНТА

А. М. Михеев, П. С. Краснощеков

Использование современных технически сложных объектов в различных научно-производственных сферах деятельности обязывает предпринимать меры по обеспечению их безопасного использования, а также мониторинга и контроля состояния. Проведение научно-исследовательских экспериментов на современных экспериментальных стендах для организации удаленного эксперимента, наиболее приближенного к реальной ситуации, является устойчивой мировой тенденцией. Об этом свидетельствуют огромное количество разработок, а также множество зарубежных и отечественных публикаций, посвященных решениям задач в области мониторинга и контроля удаленного физического объекта.

Стоимость организации современных экспериментальных стендов велика, а уровень развития современных телекоммуникационных технологий предоставляет широкие возможности, связанные с созданием распределенных информационных систем, поэтому задача создания и последующего использования систем поддержки удаленного эксперимента является весьма актуальной.

Совмещение аппаратно-программных систем с телекоммуникационными сетями и предоставление удаленного доступа к измерительному и управляющему оборудованию позволяет не зависеть от расстояния до удаленного физического объекта. Подобная интеграция двух технологий расширяет функциональные возможности систем, построенных на их основе [1].

Эффективность использования подобных систем для поддержки удаленных дистанционных экспериментов достигается благодаря возможности программным путем опираться на мощь современной компьютерной техники, при этом перестройка к изменяющимся требованиям значительно уменьшает затраты на материальное и программное обеспечение. Весьма важным является продвижение дистанционных технологий в лабораторные практикумы и в учебный эксперимент как с целью повышения эффективности, так и снижения материальных затрат на обучение в сфере инженерного образования.

Для проведения эффективных экспериментальных исследований следует использовать положения теории научного планирования эксперимента, факторного анализа, методов математической статистики, прогнозирования и поиска оптимальных решений.

Получение экспериментальных данных обеспечивает информационно-измерительная система, которая в реальных условиях имеет распределенную структуру, включающую в себя интеллектуальные многофункциональные датчики, унифицирующие измерительные преобразователи, средства хранения и цифровой обработки информации, соединенные между собой каналами связи.

Полученная измерительная информация наряду с накопленной ранее информацией и синтезированными на ее основе математическими моделями используется для идентификации состояний исследуемой сложной технической системы и протекающих в ней физических процессов и уточнения соответствующих математических моделей.

Удобным средством решения задачи идентификации состояний исследуемой системы и протекающих в ней процессов служит искусственная нейронная сеть (ИНС), а процесс ее обучения – средством хранения и уточнения соответствующих математических моделей.

Задача обучения ИНС формулируется следующим образом. Известен набор из m эталонных векторов размерностью n , описывающих состояния исследуемой сложной технической системы или протекающие в ней физические процессы [2]:

$$X_i^k = \{x_i^k : k = 0 \dots m-1; i = 0 \dots n-1\},$$

где x_i^k – элементы векторов обучающей выборки.

В процессе обучения ИНС необходимо при подаче k -го вектора

$$X_i^k = \{x_i^k : i = 0 \dots n-1\}$$

обеспечить выходной сигнал сети

$$Y = \{y_i : i = 0 \dots n-1\},$$

равный эталонному

$$Y = X^k.$$

В процессе обучения ИНС подбираются ее структура (число слоев нейронов, число нейронов в слое, их взаимосвязи и т.д.), а также параметры нейронов.

В процессе идентификации на вход ИНС подается вектор результатов измерений параметров сложной технической системы или протекающего в ней процесса, имеющий ту же размерность, что и эталонные векторы:

$$X_i^* = \{x_i^* : i = 0 \dots n-1\}.$$

Необходимо в обучающей выборке определить вектор

$$X_i^k = \{x_i^k : i = 0 \dots n-1\},$$

максимально близкий к исследуемому, и выдать его на выход ИНС

$$Y = X^k.$$

Если такой вектор определить не удастся, необходимо выдать об этом сообщение.

Архитектура ИС ПНЭ представлена на рис. 1.

Разработанная интеллектуальная информационная система поддержки научного эксперимента (ИИС ПНЭ) обеспечивает:

- сбор и обработку измерительной информации;
- обмен данными между удаленным объектом и информационной системой;
- идентификацию состояний исследуемого физического объекта по совокупности накопленной информации с помощью нейронной сети;
- визуальный контроль состояния физического объекта через WEB-интерфейс;
- дистанционное управление измерительным экспериментом.

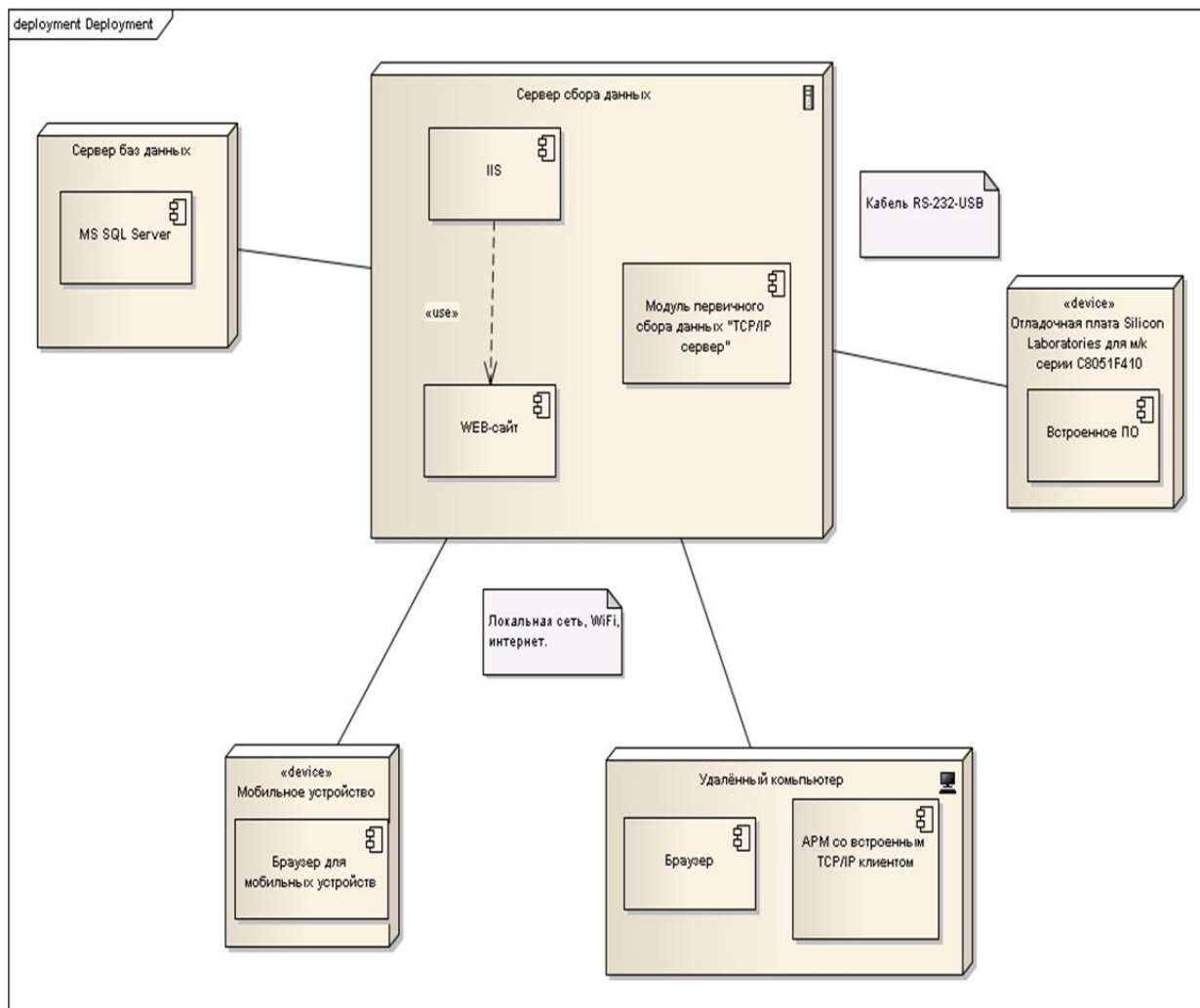


Рис 1. Архитектура ИС ПНЭ

К сожалению, ИНС относительно хорошо работают при хорошо проработанной обучающей выборке и относительно небольшом числе хорошо различающихся друг от друга ситуаций. Поэтому в настоящее время нейросетевая идентификация понимается как один из элементов технологии интеллектуального анализа данных *Data Mining* [3].

Поскольку для передачи информации используется некий несущий сигнал, выполняющий функцию транспорта для доставки информации, измерительные данные с удаленного физического объекта системы сбора и обработки данных получают в виде цифрового сигнала. Однако измерительная информация, получаемая подсистемой сбора данных за относительно короткий промежуток времени, представляет собой только первичные или так называемые сырые данные о состоянии удаленного физического объекта. Следовательно, по первичной измерительной информации, за короткий промежуток времени можно произвести только первичную оценку его состояния, а дальнейшее прогнозирование его поведения невозможно по причине отсутствия необходимого количества данных.

Data Mining – новая технология интеллектуального анализа данных с целью выявления скрытых закономерностей в виде значимых особенностей, корреляций, тенденций и шаблонов. Современные системы добычи данных используют основанные на методах искусственного интеллекта средства представления и интерпретации, что и позволяет обнаруживать растворенную в терабайтных хранилищах не очевидную, но весьма ценную информацию. В основу технологии *Data Mining* положена концепция шаблонов, отражающих фрагменты многоаспектных взаимоотношений в данных. Эти шаблоны представляют собой закономерности, свойственные выборкам данных, которые могут быть компактно выражены в понятной человеку форме. Поиск шаблонов

производится методами, не ограниченными рамками априорных предположений о структуре, выборке и виде распределений значений анализируемых показателей.

Выделяют следующие стандартные типы:

– классификацию. Наиболее распространенная задача Data Mining. Она позволяет выявить признаки, характеризующие однотипные группы объектов для того, чтобы по известным значениям этих характеристик можно было отнести новый объект к тому или иному классу;

– последовательность. Подобно ассоциациям, последовательность имеет место между событиями, но наступающими не одновременно, а с некоторым определенным разрывом во времени;

– прогнозирование. Это особая форма предсказания, которая на основе особенностей поведения текущих и исторических данных оценивает будущие значения определенных численных показателей. В задачах подобного типа наиболее часто используются традиционные методы математической статистики (в частности, регрессия), а также нейронные сети;

– анализ отклонений. Выявление наиболее нехарактерных шаблонов.

При выявлении вышеназванных закономерностей Data Mining использует большое число различных методов или их различные комбинации.

В качестве первого направления развития средств Data Mining следует выделить методы статистической обработки данных, которые можно подразделить:

– на предварительный анализ природы статистических данных (проверка гипотез стационарности, нормальности, независимости, однородности, оценка вида функции распределения и ее параметров, выделение существенных качественных факторов (дисперсионный анализ);

– выявление связей и закономерностей (корреляционный, регрессионный анализ);

– многомерный статистический анализ (кластерный, дискриминантный, факторный);

– динамические модели и прогноз на основе временных рядов.

Важное положение Data Mining – нетривиальность разыскиваемых шаблонов. Это означает, что найденные шаблоны должны отражать неочевидные, неожиданные регулярности в данных, составляющие так называемые скрытые знания. К обществу пришло понимание, что сырые данные содержат глубинный пласт знаний, при грамотной раскопке которого могут быть обнаружены настоящие самородки.

Трансформация, нормализация данных. Это необходимо для методов, которые требуют, чтобы исходные данные были в каком-то определенном виде. Дело в том, что различные алгоритмы анализа требуют специальным образом подготовленные данные, например, для прогнозирования необходимо преобразовать временной ряд с помощью скользящего окна, или вычисление агрегируемых показателей.

Для возможности реализации методов прогнозирования состояния удаленного физического необходимо накопить измерительную информацию с объекта за большой промежуток времени. Для решения данной задачи в информационной системе поддержки удаленного эксперимента в качестве хранилища данных была разработана реляционная база данных.

База данных в информационной системе поддержки удаленного эксперимента выполняет функции систематизации результатов экспериментов. Упорядоченность информации в базе данных, помимо удобств доступа, ведет к значительному сокращению аппаратных ресурсов сервера баз данных, необходимых для ее обслуживания, которая достигается путем нормализации и целостности данных.

Разработка хранилища данных, полученных с удаленного источника информации, позволяет осуществить накопление измерительной информации за большой промежуток времени для возможности дальнейшего прогнозирования состояния контролируемого объекта. Однако полезной измерительной информацией, которая возможно является необходимой для прогнозирования состояния исследуемого объекта, в хранилище данных информационной системы поддержки удаленного эксперимента является часть от всей накопленной информации. Для выявления той полезной информации, по которой возможно произвести идентификацию дальнейшего состояния исследуемого объекта, было принято решение использовать технологии предобработки и трансформации данных, а затем технологию Data Mining, выявление полезной информации в «сырых данных».

Общий принцип построения решений, использующих Data Mining, заключается в создании соответствующей модели, в ее обучении и тестировании, а также в разработке приложения, применяющего созданную модель для прогнозирования значений неизвестных атрибутов. При этом про-

цессы создания и обучения модели включают обязательное использование большого по объему набора данных, для которых известны значения атрибутов и прогнозирование которых предполагается в данной модели. Перед созданием модели, предназначенной для реальной эксплуатации, иногда создается ее прототип, обучаемый на наборе данных меньшего объема, а другой набор предназначен для тестирования созданной модели или прототипа. После подготовки исходных наборов данных выбирается алгоритм Data Mining, параметры работы алгоритма и входные поля.

В информационной системе поддержки принятия решения было принято решение использовать следующие алгоритмы, входящие в состав MS SQL Server 2008:

– алгоритм последовательной кластеризации (Microsoft Sequence Clustering) сочетает анализ последовательности операций с кластеризацией, т.е. в отличие от обычных алгоритмов кластеризации он учитывает последовательность возникновения событий. Обычно алгоритмы кластеризации используются в тех случаях, когда нет абсолютно никаких предположений о характере взаимосвязи между данными, а результаты их применения нередко являются исходными данными для других алгоритмов, например, для построения деревьев решений. Обычно они осуществляют итеративный поиск групп данных на основании заранее заданного числа кластеров. Изначально центры будущих кластеров представляют собой случайным образом выбранные точки в n -мерном пространстве возможных значений (где n – число параметров). Затем все исходные данные перебираются и в зависимости от значений параметров помещаются в тот или иной кластер, при этом постоянно происходит поиск точек, сумма расстояний которых до остальных точек в данном кластере является минимальной. Эти точки становятся центрами новых кластеров, и процедура повторяется до тех пор, пока центры и границы новых кластеров не перестанут перемещаться. Данный алгоритм далеко не всегда приводит к результату, поддающемуся логическому объяснению; он просто позволяет определить различные группы объектов или событий. Кроме того, не всегда можно с первого раза точно угадать число кластеров, отражающее реально существующее число групп. Реализация выполнена на программной платформе Microsoft SQL Server 2008 (Developer edition) с установленными аналитическими службами;

– алгоритм поиска ассоциаций (Microsoft Association) обеспечивает эффективный метод нахождения корреляций в больших наборах данных за счет поиска наборов записей, появляющихся одновременно. На основании полученных сведений генерируются правила о том, какие категории записей должны появляться вместе с наибольшей вероятностью.

В процессе исследования были проведены работы:

– разработан электронный депозитарий, обеспечивающий хранение планов экспериментов и полученной в ходе них измерительной информации:

– в качестве структуры данных электронного депозитария использована распределенная реляционная база данных, обеспечивающая дистанционный доступ к структурированной измерительной информации посредством сети передачи данных intranet и Internet.

Доступ к панели управления интеллектуальной информационной системы контроля состояния технически сложного объекта осуществляется через локальную сеть или сеть *Internet* посредством браузера. На рис. 1 представлена диаграмма развертывания, показывающая взаимосвязи между подсистемами и как осуществляется доступ к информационной системе пользователей, а также аппаратные и программные средства, необходимые для ее функционирования.

Используемые аппаратные и программные платформы обеспечивают сохранность и целостность информации при полном или частичном отключении электропитания, аварии сетей телекоммуникации, полном или частичном отказе технических средств ИИС ПНЭ.

WEB-ориентированная подсистема управления содержит слой представления, который отвечает за реакцию подсистемы на действия пользователя. Он спроектирован по шаблону проектирования *Model View Controller (MVC)* [5], который является наиболее подходящим современным шаблоном проектирования пользовательского интерфейса, в качестве программной платформы разработки была использована классическая web-ориентированная технология ASP.NET MVC и современный объектно ориентированный язык программирования C# [6, 7].

Список литературы

1. Опыт ОАО «НИИФИ» в использовании прогрессивных материалов и технологий при изготовлении интеллектуальных датчиков, микроэлектромеханических систем для систем мониторинга и контроля технически сложных объектов / А. Г. Дмитриенко, И. В. Волохов С. И. Торгашин, М. В. Тюрин // Матема-

- тическое моделирование в машино- и приборостроении : сб. науч. тр. Пенз. гос. ун-та. – Пенза, 2010. – Спецвып. № 4. – С. 90–111.
2. Имитационное моделирование процедуры нейросетевой идентификации двумерного радиолокационного сигнала / В. А. Северин, А. А. Кузнецов, М. Ю. Михеев, И. Ю. Семочкина // Вопросы радиоэлектроники. – 2009. – Т. 4, № 4. – С. 114–120.
 3. Михеев, А. М. Информационная система поддержки научного эксперимента / А. М. Михеев, С. А. Исаков // Современные информационные технологии : тр. Междунар. науч.-техн. конф. – Пенза : ПГТА, 2011. – № 14. – С. 119–122.
 4. Барсегян, А. А. Технологии анализа данных. Data Mining, Visual Mining, Text Mining, OLAP / А. А. Барсегян, М. С. Куприянов, В. В. Степаненко, И. И. Холод. – СПб. : БХВ-Петербург, 2007. – С. 384.
 5. Фаулер, М. Шаблоны корпоративных приложений / М. Фаулер. – М. : Вильямс, 2010. – С. 579.
 6. Эспозито, Д. Microsoft ASP.NET 2.0 базовый курс / Д. Эспозито. – М. : Вильямс, 2007. – С. 688.
 7. Информационно-измерительный лабораторный комплекс исследования теплопроводов электрорадиоэлементов / Н. В. Горячев, А. В. Лысенко, И. Д. Граб, Н. К. Юрков // Надежность и качество : тр. междунар. симп. : в 2 т. / под ред. Н. К. Юркова. – Пенза : Изд-во ПГУ, 2012. – Т. 2. – С. 239–241.

УДК 001.891.5

Михеев, А. М.

Алгоритмы интеллектуального анализа данных в информационной системе поддержки удаленного эксперимента / А. М. Михеев // Надежность и качество сложных систем. – 2013. – № 2. – С. 81–86.

Михеев Алексей Михайлович

аспирант, кафедра конструирования и производства радиоаппаратуры, Пензенский государственный университет
440026, г. Пенза, ул. Красная, 40.
E-mail: dagas_88@mail.ru

Краснощечков Павел Сергеевич

доктор физико-математических наук, профессор, действительный член Российской академии наук, лауреат Премии СМ СССР, главный научный сотрудник, Учреждение Российской академии наук, Вычислительный центр им. А. А. Дородницына, 119333, г. Москва, ул. Вавилова, 40.
8(499) 135-24-89

Аннотация. Рассматриваются возможности применения алгоритмов интеллектуального анализа данных в современных программно-аппаратных комплексах для поддержки научных экспериментов. Контроль объекта мониторинга осуществляется в рамках проведения научного эксперимента с помощью средств ИС ПНЭ, результаты фиксируются в базе данных.

Ключевые слова: интеллектуальные информационные системы, экспертные оценки, приобретение знаний.

A. Mikheyev

graduate student,
chair of department construction
and the production of radio equipment,
Penza State University
440026, Penza, Red street, 40.
E-mail: dagas_88@mail.ru

P. Krasnoshchekov

doctor of physico-mathematical sciences, professor,
member of the Russian Academy of Sciences,
winner of the Prize of the USSR CM,
Chief Scientific Officer,
CCAS named A. A. Dorodnitsyn,
119333, Moscow, Vavilov st., 40.
8(499) 135-24-89

Abstract. The article considers the capabilities using data mining algorithms in the modern software and hardware systems development for physical objects' state remote control. Object control is scientific experiment by means system', the results recorded in database.

Key words: intellectual information systems, expert estimates, acquisition of knowledge.