

УДК 681.518

А.В. Зивенко, магістр,
Ю.Д. Жуков, д-р техн. наук, проф.,
Нац. ун-т кораблебудування ім. адм. Макарова,
г. Николаев,
В.А. Зивенко, інженер, Николаевський нац. ун-т
ім. В.О. Сухомлинського

ФОРМИРОВАНИЕ БАЗЫ ЗНАНИЙ ПОЛИМЕТРИЧЕСКИХ СИСТЕМ НА ОСНОВЕ ПРИМЕНЕНИЯ АЛГОРИТМОВ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ

О.В. Зивенко, Ю.Д. Жуков, В.О. Зивенко. **Формування бази знань поліметричних систем на основі застосування алгоритмів інтелектуального аналізу даних.** З метою розширення функціональних можливостей комп'ютеризованих поліметричних систем вимірювання характеристик рідких середовищ розроблена структура бази знань поліметричних систем і узагальнена методика її навчання. Запропоновано алгоритм формування навчальної множини даних, що формується на об'єктах експлуатації систем. Показано, що застосування методів інтелектуального аналізу даних дозволяє отримувати нові знання для оцінки характеристик контролюваних середовищ. Розроблена база знань дозволяє застосовувати поліметричні системи в різних галузях промисловості шляхом її адаптації та перенавчання систем для оцінки характеристик принципово різних рідких середовищ.

Ключові слова: поліметричні системи; інтелектуальні системи; бази знань; інтелектуальний аналіз даних.

А.В. Зивенко, Ю.Д. Жуков, В.А. Зивенко. **Формирование базы знаний полиметрических систем на основе применения алгоритмов интеллектуального анализа данных.** С целью расширения функциональных возможностей компьютеризированных полиметрических систем измерения характеристик жидких сред разработана структура базы знаний полиметрических систем и обобщенная методика ее обучения. Предложен алгоритм формирования обучающего множества данных, получаемых на объектах эксплуатации систем. Показано, что применение методов интеллектуального анализа данных позволяет получать новые знания для оценки характеристик контролируемых сред. Разработанная база знаний позволяет применять полиметрические системы в различных отраслях промышленности путем ее адаптации и переобучения систем для оценки характеристик принципиально различных жидких сред.

Ключевые слова: полиметрические системы; интеллектуальные системы; базы знаний; интеллектуальный анализ данных.

A.V. Zivenko, Y.D. Zhukov, V.A. Zivenko. **Creation of knowledge base for polymetric systems by applying the data mining techniques.** The structure of polymetric system knowledge base and generalized training technique were developed to expand the functionality of polymetric computerized systems for measuring characteristics of liquid media. The algorithm of the training of dataset formed from the data collected during operation of industrial systems is proposed. It is shown that the application of data mining techniques allows to obtain new knowledge to estimate characteristics of a controllable media. The developed knowledge base can be adapted for different types of liquids, that allows using the polymetric systems in various industries.

Keywords: polymetric systems; intellectual systems; knowledge bases; data mining.

Неотъемлемой функциональной составляющей любой автоматизированной системы является измерение и контроль различных физических величин. Источниками измерительной информации являются измерительные преобразователи, работа которых основывается на различных физических принципах. Как правило, необходимо контролировать некоторое множество характеристик, имеющих различную природу. Для этого применяются сложные

информационно-измерительные системы (ИИС), использующие множество различных измерительных приборов, что приводит к усложнению систем обработки информационных сигналов, повышению стоимости ИИС. Среди актуальных задач, решаемых ИИС, весьма важной является задача оперативного измерения и контроля в реальном масштабе времени широкого набора характеристик разнообразных жидких сред (ЖС) техногенных процессов и производств (в различных отраслях хозяйства). Для решения данных задач все шире применяются полиметрические компьютеризированные системы (ПКС) — на основе унифицированных измерительных каналов, предназначенных для измерения нескольких различных характеристик контролируемой среды. Увеличение числа одновременно измеряемых характеристик среды может быть достигнуто за счет применения баз знаний (БЗ), содержащих алгоритмы оценки измеряемых величин по сигналам ПКС для различных видов ЖС в широком диапазоне их возможных состояний.

Недостатком применяемых сегодня на практике промышленных образцов ПКС является то, что для измерения трех и более характеристик ЖС (например, уровней раздела ЖС в емкости, их концентрации или плотности, а также их температуры) фактически используется два конструктивно совмещенных ИК, основанных на различных физических принципах действия [1, 2]. Выбор алгоритмов оценки характеристик ЖС (расчетных схем) производится на основе данных о классе исследуемой ЖС, который задается оператором. Однако, в условиях неопределенности (отсутствие информации о состоянии и классе ЖС) должна производиться идентификация ЖС, автоматический выбор подходящих расчетных схем на основании анализа доступной априорной и измерительной информации. Таким требованиям соответствуют интеллектуальные измерительные системы [3], одной из основных составляющих которых является база знаний.

Целью работы является разработка структуры базы знаний полиметрических систем для идентификации ЖС и оценки их характеристик, алгоритмов ее работы, методики сбора данных для обучения и пополнения базы соответствующими знаниями.

БЗ полиметрической системы — это особого рода база данных, предназначенная для оперирования знаниями о ЖС и интерпретации полиметрического сигнала с целью определения характеристик ЖС. БЗ содержит структурированную информацию о правилах и алгоритмах расчета показателей ЖС по данным полиметрического сигнала. БЗ полиметрической системы может содержать алгоритмические (процедурные) знания [4] — алгоритмы (программы, процедуры), вычисляющие функции, выполняющие преобразования, решающие точно определенные задачи. Процедурные знания закладываются экспертом на этапе обучения БЗ. Каждому алгоритму, как единице знаний ставится в соответствие некоторое множество правил его применения.

Библиотека алгоритмов должна содержать набор алгоритмов для идентификации ЖС, оценки характеристик ЖС, а также текстовые описания алгоритмических действий — процедурные знания с описанием. Кроме того библиотека алгоритмов должна содержать подробное описание всех выходных параметров, передаваемых в следующий этап вычислительного процесса — шаблон результата для непосредственной визуализации пользователю. Структура разработанной базы знаний имеет вид, изображенный на рис. 1.

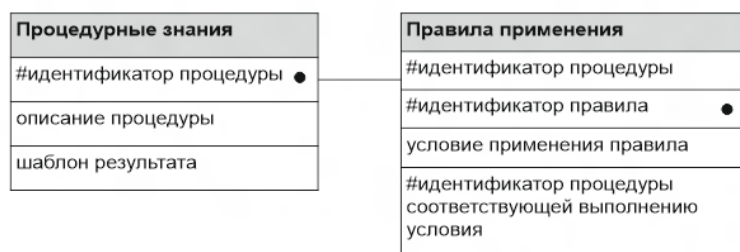


Рис. 1. Структура базы знаний полиметрической системы

Каждый алгоритм в результате своей работы генерирует новый объект, который содержит не только результаты своей работы, но и всю исходную информацию и передает ее в соответствии с правилами на следующий этап обработки. Выбор следующего этапа зависит от результата обработки на данном этапе, а также правил, записанных в таблице правил.

Таким образом, каждому алгоритму (расчетной схеме, классификатору), а точнее результату его работы, соответствует набор правил, определяющих последовательность дальнейших действий — поиска подкласса или непосредственного вычисления компонент вектора состояния ЖС. Вначале работы БЗ запускается процедура, являющаяся корневой — процедура базовой идентификации ЖС. Результатом выполнения данной процедуры является идентификатор базового класса ЖС. После окончания корневой процедуры производится выборка множества правил, соответствующих данной процедуре и происходит проверка соответствия результата выполнения данной процедуры, условиям, записанным в правилах. Множество правил соответствующих результату выполнения функции на текущем этапе вычислений образует полную группу событий. На языке алгебры предикатов [5] данное условие можно сформулировать следующим образом. Для этого введем переменную — результат классификации ЖС — *class*, которая принимает одно из значений: c_1 — жидкий нефтепродукт, c_2 — спиртоводный раствор, c_3 — сжиженный газ, c_4 — не определено. Тогда область определения переменной *class* задается следующим уравнением:

$$class^{c_1} \vee class^{c_2} \vee class^{c_3} \vee class^{c_4} = 1. \quad (1)$$

Аналогично (1) задаются области определений других переменных, которые являются результатом выполнения различных процедур БЗ.

Правила применения процедурных знаний также могут быть заданы на языке алгебры предикатов. Например, если в результате вычисления определен класс ЖС $class^{c_1}$ и подкласс ЖС $sub_class^{c_81}$, то для последующих вычислений должна использоваться процедура, соответствующая данному подклассу ЖС — p^{c_81} , что может быть описано следующим образом:

$$class^{c_1} \wedge sub_class^{c_81} \supset p^{c_81}.$$

Результат выполнения каждой процедуры вычислений (классификации, поиска и т.д.) заканчивается проверкой условий, записанных в базе правил и соответствующих данной процедуре. Терминальные процедуры БЗ производят оценку требуемых параметров и передают в модуль визуализации не только вектор состояния ЖС, но и логическую цепь получения данных значений.

Структура разработанной БЗ представляет собой гибридную модель: правилами задается жесткая логика принятия решений в зависимости от результата выполнения процедур (применение продукций), а процедуры могут включать «мягкие» методы вычислений.

Процесс обучения системы в данном случае заключается в выявлении связей параметров полиметрического сигнала и характеристик ЖС, формализации закономерностей в виде расчетных схем и формировании правил их применения (приобретении знаний), которые могут быть получены различными путями (использование существующей справочной информации по конкретным ЖС, экспертные оценки, результаты экспериментальных исследований и опытной эксплуатации). Одним из способов получения расчетных схем и формирования правил их применения является использование современных средств извлечения знаний из данных [6, 7]. Данными в этом случае является совокупность информации о состояниях ЖС и о параметрах полиметрических гиперсигналов ПКС [8], соответствующих этим состояниям ЖС.

Так, нейронные сети могут быть успешно использованы для оперативной идентификации класса ЖС по их электрофизическим параметрам. Под идентификацией в данном случае понимается вывод о принадлежности ЖС некоторому классу ЖС на основе сопоставления их свойств. В результате экспериментов были накоплены данные о частотных спектрах диэлектрической проницаемости растворов различных спиртов находящихся в различных состояниях (при различных температурах и концентрациях компонент) [9]. На основе

накопленного массива данных была обучена нейронная сеть, которая позволяет определить класс исследуемого раствора спирта по значениям диэлектрической проницаемости на различных частотах. Полученная нейронная сеть правильно идентифицирует класс ЖС более чем в 99 % случаев. Оперативная идентификация класса ЖС позволяет учитывать данные о классе ЖС при обработке полиметрического сигнала и за счет этого повысить точность одновременной оценки множества характеристик ЖС.

Применение кластерного анализа позволяет выявлять группы близких по составу ЖС и формировать «свои» расчетные схемы для каждой выявленной в исходном множестве группы, а также выявлять условия принадлежности ЖС к той или иной группе. Такой подход позволяет повысить точность оценки характеристик ЖС. Например, указанный подход позволил повысить точность оценки октанового числа бензинов различных марок до уровня приемлемого при оперативном коммерческом контроле качества топлив [10].

Обобщенная методика обучения ПКС включает следующие этапы:

- формирование обучающего множества данных (сигналов), получаемых для разных ЖС с различными известными сочетаниями значений оцениваемых характеристик ЖС;
- подготовка обучающего множества данных — очистка данных от различных помех, мешающих их корректному анализу, статистическая обработка данных;
- обработка полученных экспериментально полиметрических сигналов;
- поиск корреляций вычисленных параметров и показателей качества ЖС, выявление взаимосвязей параметров сигнала и характеристик ЖС, выбор информативных параметров для построения расчетных схем;
- построение и анализ расчетных схем для оценки количества и характеристик качества ЖС. Тестирование построенных расчетных схем на опытных данных, не вошедших в обучающую выборку;
- внедрение расчетных схем на производстве и проверка адекватности предложенных решений в реальных условиях эксплуатации.

Одним из наиболее важных этапов при обучении ПКС является формирование обучающей выборки — длительный и трудоемкий процесс. Трудоемкость процесса обусловлена требованием воспроизведения большого числа различных возможных состояний ЖС с фиксацией условий получения полиметрического сигнала. Одним из внедряемых способов преодоления трудности получения обучающих выборок является использование промышленных полиметрических систем в режиме сбора и накопления обучающих данных.

В режиме сбора данных ПКС собирает сигналы системы, соответствующие текущему состоянию ЖС. Данные о текущем состоянии ЖС могут быть получены в лабораториях объектов установки ПКС, где ведется учет и постоянный мониторинг показателей ЖС (например, лаборатории нефтеперерабатывающих предприятий, комплексов по перевалке жидких грузов, предприятий пищевой промышленности и т.п.).

Данные о состоянии ЖС записываются в базу данных экспериментов. Запись может производиться в одном из двух режимов. В ручном режиме, при участии оператора системы «САДКО-М», с заданной периодичностью в базу данных заносятся текущие характеристики ЖС по данным лаборатории и соответствующий полиметрический сигнал.

Если предприятие, эксплуатирующее ПКС, использует автоматизированную систему управления технологическим процессом (АСУТП), которая получает информацию от разнообразных средств измерительной техники (СИТ), то процесс сбора данных о ЖС может производиться в автоматическом режиме.

АСУТП получает информацию о состоянии ЖС (в приведенном случае бензинах различных марок — рис. 2) от ряда различных систем и контрольных приборов (помимо ПКС), что и позволяет накопить информацию для последующего анализа полиметрических сигналов для контролируемых ЖС, а также мониторинга работы ПКС на объекте ее эксплуатации.

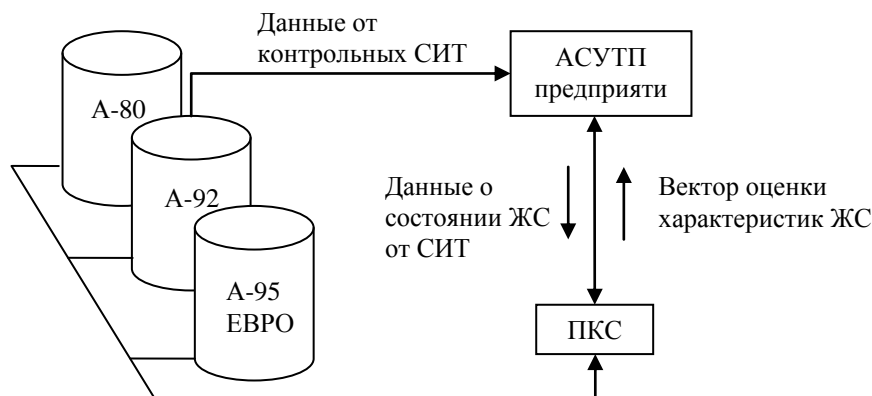


Рис. 2. Схема интеграции ПКС в АСУТП для выдачи вектора оценки состояния ЖС и сбора обучающих данных

Данная методика и соответствующее программное обеспечение приняты в опытную эксплуатацию на предприятии-изготовителе ПКС «САДКО™» - научно-производственном объединении «АМИКО» (г. Николаев).

База данных, формируемая на объекте эксплуатации, может также использоваться в качестве журнала, в котором отражается история технологического процесса. Особый интерес эта информация представляет для изучения медленно-изменяющихся свойств ЖС.

Накопленные данные периодически передаются в центр сервисного обслуживания систем — подразделение предприятия-изготовителя НПО «АМИКО», где анализируются специалистами с целью получения новых знаний и профилактических мероприятий.

Разработана общая структура базы знаний ПКС измерения характеристик ЖС, методика сбора данных и последующего обучения ПКС. Показано, что на основе баз данных, содержащих сигналы и сведения о соответствующем состоянии ЖС, с помощью методов интеллектуального анализа данных могут быть получены знания, позволяющие оценивать множество характеристик ЖС по одному полиметрическому сигналу. Разработанная БЗ, позволяет применять ПКС в различных отраслях промышленности путем ее адаптации и переобучения ПКС для оценки характеристик принципиально различных ЖС, например, спиртоводных растворов, топлив, сжиженных газов и т.д.

Литература

1. Гордеев, Б.Н. Развитие теории и практическое применение компьютеризированных полиметрических систем оперативного контроля количественных и качественных характеристик жидких сред (энергосистем): дис. д-ра техн. наук: 05.13.05: защищена 29.03.2011: утв. 11.11.2011. — К., 2011. — 422 с.
2. Гордеев, Б.Н. Полиметрический датчик / Б.Н. Гордеев, Ю.Д. Жуков, А.В. Зивенко, В.А. Зивенко // Тези IV Міжнародної науково-технічної конференції «Датчики, прилади та системи — 2008». — Черкаси-Гурзуф, 2008. — С. 63 — 64.
3. Раннев, Г.Г. Интеллектуальные средства измерений : учебник для студ. высш. учеб. заведений / Г.Г. Раннев. — М.: Издательский центр «Академия», 2010. — 272 с.
4. Костюк, А.И. Базы данных и знаний / А.И. Костюк. — Таганрог: Изд-во ТРТУ, 1999. — 175 с.
5. Шабанов-Кушнарченко, Ю.П. Теория интеллекта. Математические средства / Ю.П. Шабанов-Кушнарченко. — Харьков: Вища школа, 1984. — 144 с.
6. Чубукова, И.А. Data Mining: учебное пособие / И.А. Чубукова. — М.: Интернет-ун-т информац. технол.: БИНОМ: Лаборатория знаний, 2006. — 382 с.
7. Witten, I.H. Data Mining: Practical Machine Learning Tools and Techniques, Third Edition (The Morgan Kaufmann Series in Data Management Systems) / I.H. Witten, E. Frank, M.A. Hall. — Morgan Kaufmann, 3 edition, 2011. — 664 p.

8. Жуков, Ю.Д. Теория полиметрических измерений / Ю.Д. Жуков // Материалы МНТК «Инновации в судостроении и океанотехнике». — Николаев: НУК, 2010. — С. 22 — 23.
9. Зивенко, А. В. Математическая модель полиметрии спиртоводных растворов [Электронный ресурс] // А.В. Зивенко, Ю.Д. Жуков / Электронне видання «Вісник Національного університету кораблебудування». — 2011. — № 5. — <http://ev.nuos.edu.ua>. — 18.12.12.
10. Зивенко, А.В. Оперативная полиметрическая оценка октанового числа бензина в рамках информационно—управляющей системы «SADKO—OIL» / А.В. Зивенко, Ю.Д. Жуков // Проблемы автоматизации та електрообладнання транспортних засобів: Матеріали всеукраїнської науково-технічної конференції з міжнародною участю. — Миколаїв: НУК, 2012. — С. 24 — 27.

References

1. Gordeev, B.N. Razvitie teorii i prakticheskoe primeneniye komp'yuterizirovannykh polimetricheskikh sistem operativnogo kontrolja kolichestvennykh i kachestvennykh harakteristik zhidkikh sred (energonositelej) [Development of the theory and practical application of the computer-aided polymetric systems for on-line monitoring of quantitative and qualitative characteristics of liquid media (energy carriers)]: doctoral thesis on specialty 05.13.05: defended 29.03.2011; approved: 11.11.2011 — Kiev, 2011. — 422 p.
2. Gordeev, B.N. Polimetricheskij datchik [Polymetric sensor] / B.N. Gordeev, Ju.D. Zhukov, A.V. Zivenko, V.A. Zivenko // Proceedings of the IV international scientific and technical conference «Sensors, devices and systems — 2008». — Cherkasi-Gurzuf, 2008. — pp. 63 — 64.
3. Rannev, G.G. Intel'ektual'nye sredstva izmerenij: uchebnyj dlja stud. vyssh. ucheb. zavedenij [Intelligent measuring instruments: a textbook for university students] / G.G. Rannev. — Moscow, 2010. — 272 p.
4. Kostjuk, A.I. Bazy dannyh i znaniy: Kurs lekcij [Data and Knowledge Bases: Lectures] / Kostjuk A. I. — Taganrog, 1999. — 175 p.
5. Shabanov-Kushnarenko, Ju. P. Teorija intellekta. Matematicheskie sredstva [Theory of Intelligence. Mathematical Means] — Kharkov, 1984. — 144 p.
6. Chubukova, I.A. Data Mining: uchebnoe posobie [Data Mining: Textbook] / Moscow, 2006. — 382 p.
7. Witten, I.H. Data Mining: Practical Machine Learning Tools and Techniques, Third Edition (The Morgan Kaufmann Series in Data Management Systems) [Text] / I.H. Witten, E. Frank, M.A. Hall. — Morgan Kaufmann, 3 edition, 2011. — 664 p.
8. Zhukov, Yu.D. Teorija polimetricheskikh izmerenij [Theory of Polymetric Measurements] / Yu.D. Zhukov // Proceedings of ISTC «Innovations in Shipbuilding and Ocean Engineering». — Nikolaev, 2010. — pp. 22 — 23.
9. Zivenko, A. V. Matematicheskaja model' polimetrii spirtovodnyh rastvorov [Elektronnyy resurs] [Mathematical model of polymetry of alcohol-water solutions [Electronic resource]] / A.V. Zivenko, Ju.D. Zhukov // NUS Journal. Electronic Edition. — 2011. — # 5. — Available at <http://ev.nuos.edu.ua>. — 18.12.12.
10. Zivenko, A.V. Operativnaja polimetricheskaja ocenka oktanovogo chisla benzina v ramkah informacionno—upravljajushhej sistemy «SADKO—OIL» [Operative polymetric evaluation of octane in the information-control system «SADKO—OIL»] / A. V. Zivenko, Ju. D. Zhukov // Problems of Automation and Electric Vehicles: Proceedings of the AllUkrainian scientific conference with international participation. — Nikolaev, 2012. — pp. 24 — 27.

Рецензент д-р техн. наук, проф. Одес. нац. политехн. ун-та Крисилов В.А.

Поступила в редакцию 19 декабря 2012 г.