УДК 004.942

А.А. Голубева

Интеллектуальный анализ пространственно-временных данных объектов инженерной инфраструктуры

Предложен подход к разработке методов автоматизированной диагностики, основанный на применении теории нечеткой логики и эвристических методов. Данный подход заключается в применении гибридных методов и алгоритмов, в основе которых лежат метаэвристики и численные методы.

Ключевые слова: функция принадлежности, тип нечеткой системы, база правил.

Мониторинг и анализ состояния объектов инженерной инфраструктуры являются важным этапом жизненного цикла инженерных коммуникаций. Существующие методы мониторинга и анализа базируются в основном на визуальном обследовании коммуникаций и рассчитаны на проведение больших организационных мероприятий, требуют постоянного привлечения значительных трудовых и денежных ресурсов. И в то же время они не предполагают использование компьютерных технологий [1].

Современные инженерные коммуникации характеризуются наличием информации, которую невозможно получить непосредственно от первоисточников – подземных объектов в реальный отрезок времени из-за необходимости проведения долгосрочных работ. Тем не менее при решении задач обеспечения эффективной эксплуатации и функционирования инженерных коммуникаций требуется учет и такой информации, которой присуща некоторая неопределенность. Учет априорной информации о характеристиках инженерных коммуникаций и условиях их эксплуатации позволил бы принимать точные решения по оценке состояния. Однако в реальной жизни информация об условиях эксплуатации инженерных коммуникаций обычно является неполной или отсутствует вообще.

Проблема принятия решений в условиях неопределенности занимает важное место в общей теории принятия решении. Успешное решение данной проблемы в настоящее время невозможно без применения новых информационных технологий, составной частью которых являются интеллектуальные средства обработки информации. Для описания неопределенности современная теория принятия решений широко применяет, в частности, аппарат нечетких множеств.

Традиционный подход к проблеме принятия решений основан на использовании классических методов многокритериального анализа и предполагает разработку и создание сложных, зачастую многоуровневых систем поддержки принятия решений, базирующихся на математических моделях, обеспечивающих учет большого количества параметров и критериев, и характеризуется значительными вычислительными затратами и высокой стоимостью разработки.

Задачи принятия решений в условиях неопределенности представляют собой слабоструктурированные или неструктурированные задачи. Применение теории нечетких множеств и её приложений позволяет построить формальные схемы решения задач, характеризующихся той или иной степенью неопределенности, которая может быть обусловлена неполнотой, внутренней противоречивостью, неоднозначностью и размытостью исходных данных, представляющих собой приближенные количественные или качественные оценки параметров объектов.

Неполнота и неточность информации могут заключаться: в принципиальной невозможности полного сбора и учета информации или ее экономической нецелесообразности об анализируемом объекте или процессе; в некоторой недостоверности и недостаточности исходной информации об анализируемом объекте или процессе; в возможности проявления таких свойств анализируемого объекта или процесса, существование которых не предполагалось. Кроме того, неточность, неполнота и неопределенность исходных данных могут быть вызваны недостаточными знаниями экспертов специфики конкретной прикладной задачи.

Для решения описанных проблем возникает необходимость в разработке программных средств, использующих для классификации и аппроксимации пространственно-временных данных с учетом возможной неполноты и/или неточности информации. Одной из наиболее важных задач здесь является разработка методов автоматизированной диагностики (идентификации) неисправностей и общего состояния инженерных коммуникаций, что позволило бы использовать результаты такой диагностики в автоматизированных системах принятия решений.

Предлагаемый подход к разработке методов автоматизированной диагностики основан на применении теории нечеткой логики и эвристических методов и заключается в применении гибридных методов и алгоритмов, в основе которых лежат метаэвристики и численные методы. Метаэвристики можно использовать в качестве первичной грубой настройки нечеткой системы, а численные методы – в качестве последующей тонкой настройки.

Описанный подход позволяет создавать адаптивные модели высокой точности, обеспечивающие достоверные прогнозы в сложных инженерно-технических системах.

В данной статье в качестве примера решения проблемы обработки информации с учетом неполноты и неточности была выбрана задача аппроксимации пространственно-временных данных, описывающих состояние элементов инженерной инфраструктуры.

Суть задачи аппроксимации данных при помощи методов интеллектуального анализа данных заключается в следующем: существует неполная информация об объектах инженерных коммуникаций. На основании существующей информации необходимо спрогнозировать неизвестные параметры исследуемых объектов.

Предлагаемый метод решения задач диагностики городских инженерных коммуникаций предоставляет возможность работать с априорной информацией о состоянии городских инженерных коммуникаций в условиях неопределенности и неполноты.

Традиционные методы восстановления данных достаточно сложны и требуют многочисленных людских и материальных затрат. В связи с этим решение задачи диагностики неисправностей инженерных коммуникаций на основе традиционного подхода оказывается затруднительным и возникает необходимость разработки новых подходов – алгоритмов и систем принятия решений в условиях неопределенности и неполной информации.

Важным этапом диагностики инженерных коммуникаций является сбор априорной информации о характеристиках инженерных коммуникаций. Важно помнить, что инженерные коммуникации состоят из множества сетей, которые имеют разные характеристики. Так, например, в определенном территориально отделенном районе, могут пролегать водопроводные, канализационные сети и др. Соответственно, параметры сетей, такие как толщина труб, материал, глубина залегания и др., будут отличаться.

После того как априорная информация была собрана, необходимо определить входные и выходные параметры системы, которыми являются характеристики объектов, задав соответствующее количество термов и выбрав подходящую функцию принадлежности.

Например, существует таблица с данными, строка которой характеризует объект, а столбец – свойство. Свойства объектов будем называть параметрами, такими как «возраст трубы», «качество строительных работ», «качество ремонтных работ», «состояние арматуры», «отложение солей на трубах», «качество земли и грунтовых вод», «коррозия трубы», «утечка воды», «состояние арматуры». А в свою очередь под объектом будем понимать непосредственно сами объекты сети.

На данном этапе развития науки практически невозможно заранее точно сказать, как повлияет на результат вычислений выбор функций принадлежности и количество термов. Поэтому чтобы прийти к наиболее приемлемому выводу, необходимо проводить эксперименты и наблюдать за результатами вычислений. Выбрав удовлетворяющую ошибку вычислений (усредненное отклонение результатов расчета от реальных данных, на которых проводится обучение), можно определить наиболее подходящую функцию принадлежности.

Функция принадлежности может быть представлена одним из следующих типов:

– Треугольный – рационально использовать при необходимости быстрой классификации/апроксимации и отсутствии требований высокого качества (необходим экспресс-анализ системы), т.е. пользователю важна в первую очередь скорость обработки данных.

– Гауссовское распределение (функция имеет гладкий вид) – если важна устойчивость распределения и есть необходимость уточнения данных экспресс-анализа. Функция Гаусса является оптимальной и дает неплохие результаты по скорости обучения, скорости классификации и по эффективности классификации.

– Трапеция (функция имеет линейный и местами ступенчатый вид) – использование трапециевидной функции обусловлено наличием специфической выборки и отсутствием жестких ограничений по времени.

– Парабола – параболическая функция принадлежности является наилучшей по скорости обучения. Эффективность классификации также находится на высоком уровне.

После инициализации переменных следует произвести настройку антецедентов и консеквентов правил, предварительно определив систему нечеткого вывода и построив базу правил. Были рас-

смотрены следующие типы нечетких систем для задач аппроксимации данных: Синглтон, Мамдани, Такаги–Сугено.

По данным множества научных экспериментов было эмпирически установлено, что выбор системы нечеткого вывода согласуется со значением коэффициента регрессии следующим образом:

1) если коэффициент регрессии более 0,7, то следует использовать систему нечеткого вывода типа Такаги–Сугено;

2) если коэффициент регрессии не превышает 0,2, то следует использовать систему нечеткого вывода типа Мамдани;

3) если коэффициент от 0,2 до 0,7, то следует использовать систему нечеткого вывода типа Синглтон.

Используемые системы нечеткого вывода работают с множеством входов, но только одним выходом в каждый момент времени.

В теории нечеткой логики встречается понятие «лингвистическая переменная» (ЛП). Данное понятие позволяет адекватно отразить приблизительное словесное описание некоторых характеристик (параметров объектов) и состояния инженерных коммуникаций в тех случаях, когда точное описание либо отсутствует, либо является слишком сложным, либо требует больших временных и финансовых затрат. Описание предметной области может быть проведено при помощи лингвистических переменных и правил естественного языка.

Основой для описания предметной области является нечеткое высказывание вида:

$$x_i \text{ есть } A_{ij} \text{ или } x_i = A_{ij},$$

где x_i – имя i-й лингвистической переменной (параметра объекта); A_{ij} – j-элемент терм множества i-й лингвистической переменной из представленной области.

Отображение вход/выход может быть представлено как множество нечетких правил типа «ЕСЛИ-ТО». Каждое правило состоит из двух частей: условной и заключительной. Антецедент или условная часть (ЕСЛИ-часть) содержит утверждение относительно значений входных переменных, в консеквенте или заключительной части (ТО-части) указывается значение, которое принимает выходная переменная. Таким образом, нечеткая система типа «много входов – один выход» может быть задана нечеткими правилами следующего вида:

$$\text{правило 1: ЕСЛИ } x_1 = A_{11} \text{ И } x_2 = A_{21} \text{ И} \ldots \text{ И } x_m = A_{m1} \text{ ТО } r = R_1;$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$\text{правило } n\text{: ЕСЛИ } x_1 = A_{1n} \text{ И } x_2 = A_{2n} \text{ И} \ldots \text{ И } x_m = A_{mn} \text{ ТО } r = R_n,$$

где x_1, x_2, …, x_m – входные переменные (известные параметры объекта); r – выходная переменная (неизвестный/пропущенный параметр объекта); A_{ij} – нечеткие области определения входных переменных; R_n – значение выходной переменной, которое может быть представлено как действительное число (Синглтон), либо как функция (Такаги–Сугено), определенная на входных переменных, либо как нечеткая область определения выходной переменной (Мамдани). Каждая нечеткая область A_{ij} связана с функцией принадлежности $\mu A_{ij}(x_i)$.

Наиболее популярны три основных типа нечетких систем вида «много входов – один выход».

Нечеткая модель типа Синглтон задается правилами вида:

$$\text{правило } i\text{: ЕСЛИ } x_1 = A_{1i} \text{ И } x_2 = A_{2i} \text{ И} \ldots \text{ И } x_m = A_{mi} \text{ ТО } r = a_i,$$

где a_i – действительное число.

Модель типа Такаги–Сугено имеет правила следующего вида:

$$\text{Правило } i\text{: ЕСЛИ } x_1 = A_{1i} \text{ И } x_2 = A_{2i} \text{ И} \ldots \text{ И } x_m = A_{mi} \text{ ТО } r = f_i(x_1,\ldots,x_m),$$

где f_i – линейная функция, определенная на переменных x_1, x_2, …, x_m.

Модель типа Мамдани задается правилами следующего вида:

$$\text{правило } i\text{: ЕСЛИ } x_1 = A_{1i} \text{ И } x_2 = A_{2i} \text{ И} \ldots \text{ И } x_m = A_{mi} \text{ ТО } r = B_i,$$

где B_i – терм лингвистической переменной.

Более подробно о системах нечеткого вывода можно ознакомиться в статье [2].

Для настройки базы правил, описанной выше, применяется ряд методов: метод наименьших квадратов (МНК), метод роящихся частиц (РЧ), гибридный алгоритм (настройка методом роящихся частиц, последующая настройка методом наименьших квадратов).

При выборе метода настройки как и при выборе функции принадлежности, невозможно заранее определить, какой метод даст наилучший результат. Определить наилучший метод с точки зрения получаемых результатов можно только на этапе практического обучения системы на конкретных данных.

Метод РЧ является метаэвристическим методом и часто является наиболее успешно применяемой эвристикой для обучения нечетких систем, так как данный метод защищен от «застревания» в локальных решениях. То есть если известно, что область решения имеет много локальных не лучших решений, наилучшим с точки зрения практического применения является метод РЧ [2].

МНК является классическим алгоритмом оптимизации и находит оптимальное решение за одну итерацию, поэтому он является менее затратным по ресурсам, чем аналогичные методы, при достижении той же точности. Рекомендацией применения МНК является необходимость получить хорошее и быстрое решение [2].

В случае если нам необходимо улучшить решение, полученное другими методами, либо изначально предъявляются высокие требования к точности результатов, рекомендуется применение гибридного алгоритма. Его использование позволит объединить преимущества метаэвристических методов с преимуществами классических методов, основанных на производных. Недостатком данного метода являются большие временные затраты.

Одним из заключительных этапов процесса настройки обучения нечеткой системы является выбор типа нечеткой системы, так называемого алгоритма вывода.

После того как подобрали все необходимые параметры и методы, произвели настройку системы, можно запустить процесс аппроксимации данных.

Выполнение данной работы проводилось при финансовой поддержке Министерства образования и науки Российской Федерации в рамках мероприятия 2.4 федеральной целевой программы «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2007–2013 годы», проект «Разработка Web-ориентированных геоинформационных технологий формирования и мониторинга электронного генерального плана инженерной инфраструктуры», государственный контракт № 07.524.11.4013 от 03 ноября 2011 г.

Литература

1. Гриценко Ю.Б. Программное обеспечение интеллектуального анализа пространственно временных данных объектов инженерной инфраструктуры / Ю.Б. Гриценко, А.А. Голубева // Aktualne problemy nowoczesnych nauk – 2012: Materiały viii międzynarodowej naukowi-praktycznej konferencji, 07-15 czerwca 2012 roku. – Przemysl: Nauka I studia, 2012. – Vol. 47. – С. 89–92.

2. Ходашинский И.А. Основанные на производных и метаэвристические методы идентификации параметров нечетких моделей / И.А. Ходашинский, В.Ю. Гнездилова, П.А. Дудин, А.В. Лавыгина // Тр. VIII Междунар. конф. «Идентификация систем и задачи управления» SICPRO '08. – М.: Институт проблем управления им. В.А. Трапезникова РАН. – 2009. – С. 501–529.

Голубева Александра Александровна
Аспирант каф. АОИ ТУСУРа
Тел.: 8-913-825-84-26
Эл. почта: Sasha_karateka@mail.ru

Golubeva A.A.

Mining spatio-temporal data of the engineering infrastructure

In the paper we offer an approach to the development of methods for the automated diagnosis based on the theory of fuzzy logic and heuristic methods. This approach uses hybrid methods and algorithms, which are based on meta heuristic and numerical methods.

Keywords: membership function, the type of fuzzy system, rule base.