

Л. Ю. Емалетдинова, А. С. Катасёв, А. П. Кирпичников

НЕЙРОНЕЧЕТКАЯ МОДЕЛЬ АППРОКСИМАЦИИ СЛОЖНЫХ ОБЪЕКТОВ С ДИСКРЕТНЫМ ВЫХОДОМ

Ключевые слова: моделирование, аппроксимация, сложный объект, нечетко-продукционное правило, нейронечеткая модель, база знаний.

В работе решается задача аппроксимации сложных объектов с дискретным выходом на основе информационного подхода к моделированию. Предлагается вид нечетких правил и алгоритм логического вывода на правилах. Описывается нейронечеткая модель для формирования базы знаний. Производится аппроксимация известных наборов данных и сравнение полученных результатов с результатами других авторов. На примере формирования баз знаний экспертных диагностических систем в медицине, нефтяной отрасли и информационной безопасности показывается эффективность предложенного подхода.

Keywords: modeling, approximation, complex object, fuzzy production rule, neuro-fuzzy model, knowledge base.

This paper solves the task of complex objects approximation with digital output based on information approach to modeling. Propose a model of fuzzy rules and the inference algorithm on the rules. Describes the neuro-fuzzy model for a knowledge base generation. The approximation of known data sets and comparison of the results with results of other authors is performed. Using the example of knowledge bases generation of the expert diagnostic systems in medicine, oil industry and information security shows effectiveness of the proposed approach.

Моделирование является универсальным способом познания окружающего мира, позволяющим изучать свойства и закономерности сложных объектов путем построения их упрощенных аналогов – моделей. В настоящее время выделяют аналитический и информационный подходы к моделированию [10]. В первом случае выбирается готовая аналитическая модель таким образом, чтобы она адекватно отражала реальность. Во втором случае модель строится на основе анализа накопленных данных, учитывая специфику моделируемого объекта [8,14]. Очевидно, что информационный подход более предпочтителен при моделировании сложных объектов, для которых не существует готовой аналитической модели или ее построение затруднительно.

Сложные объекты, как правило, характеризуются следующими особенностями [9]:

- наличием большого числа входов, выходов и состояний, полный учет которых либо нереален, либо резко увеличивает сложность модели;
- разнотипностью описывающих параметров и нелинейным характером внутренних связей и закономерностей;
- неопределенностью, нечеткостью и неполнотой данных на входе, выходе объекта и в переходе между его состояниями;
- плохой структурированностью.

Кроме того, многообразие и сложность процессов, протекающих в реальном мире, приводят к тому, что все они в той или иной степени оказывают свое воздействие на моделируемые объекты, что еще более усложняет их. Это требует применения соответствующего математического аппарата и накладывает ограничения на квалификацию специалистов, занимающихся моделированием.

Традиционные подходы, основанные на аппарате математической статистики или имитационном моделировании, не позволяют строить адекватные модели сложных объектов в условиях ограни-

ченности вычислительных и временных ресурсов. Для решения данной проблемы в настоящее время прибегают к методам искусственного интеллекта на основе нечеткой логики, позволяющей описывать сложные объекты на естественном языке при помощи лингвистических переменных и алгоритмов нечеткого логического вывода.

В реальных практических задачах системы нечеткого логического вывода применяются для аппроксимации сложных объектов, как с непрерывными, так и с дискретными выходами [15]. Объекты с непрерывным выходом соответствуют задачам аппроксимации гладких функций, возникающим в прогнозировании, многокритериальном анализе, управлении техническими объектами и т.д. Объекты с дискретным выходом соответствуют задачам классификации в медицинской и технической диагностике, в распознавании образов, в ситуационном управлении и принятии решений в других областях.

Рассмотрим постановку задачи нечеткой аппроксимации объектов с дискретным выходом [13]. Пусть для некоторого объекта имеется множество входных переменных $X = (x_1, x_2, \dots, x_n)$, а также множество дискретных классов решений $D = (d_1, d_2, \dots, d_m)$, соответствующих выходной переменной y . Пусть также имеется база знаний, состоящая из нечетко-продукционных правил, в которых определены функции принадлежности, позволяющие представлять переменные x_i , $i = \overline{1, n}$ в виде нечетких множеств. Требуется фиксированному вектору входных переменных $X^* = (x_1^*, x_2^*, \dots, x_n^*)$ поставить в соответствие конкретное решение $y \in D$ (см. рис. 1).

Для решения данной задачи необходима разработка вида нечетко-продукционных правил, алгоритма логического вывода на правила, а также инструментального средства автоматизированного формирования базы знаний. На основе анализа мо-

делей представления знаний и с учетом особенностей обрабатываемых данных предлагается нечетко-продукционная модель следующего вида [6]:

$$\begin{aligned} &\langle \text{ЕСЛИ } x_1 = \bar{A}_1(w_1) \text{ И } \dots x_n = \bar{A}_n(w_n) \\ &\text{ТО } y = B \rangle [CF], \end{aligned} \quad (1)$$

где $x_i, i = \overline{1, n}$ – входные переменные, $w_i \in [0, 1]$ – веса условий, \bar{A}_i – ограничения для четких и нечетких входных условий, y – выходная переменная, B – четкое значение выхода, $CF \in [0, 1]$ – достоверность правила.

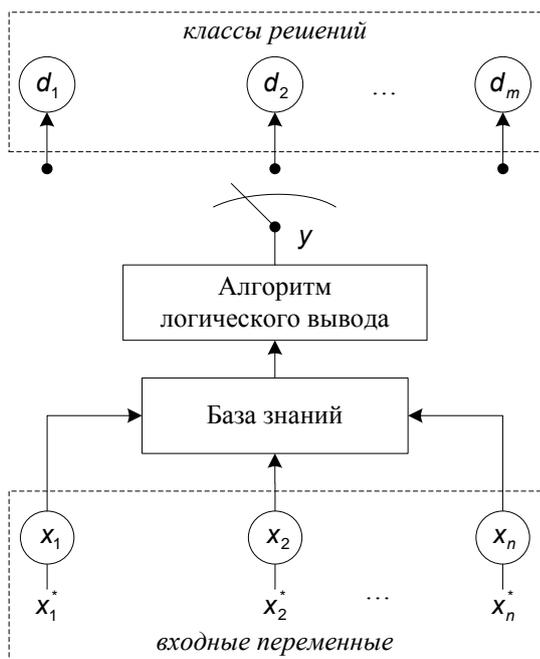


Рис. 1 - Схема аппроксимации объекта с дискретным выходом

Особенность правил вида (1) заключается в возможности применения разнотипных входных и выходных параметров, обработки четких и нечетких входных данных, учета весов условий в правиле и достоверности каждого правила.

Для организации логического вывода на правилах вида (1) введем следующие обозначения:

- $R \in [0, 1]$ – степень выполнения условной части правила $R = \min(\mu_{\bar{A}_i}(x_i^*))$, где $x_i^*, i = \overline{1, n}$ – четкие значения n входных параметров правила, $\mu_{\bar{A}_i}(x_i^*)$ – степени принадлежности входных значений x_i^* к \bar{A}_i ;

- $T \in [0, 1]$ – степень доверия к решению правила $T = \frac{\sum_{kn=1}^{n_{kn}} w_{kn}}{\sum_{i=1}^n w_i}$, где $w_i, i = \overline{1, n}$ – веса всех ограничений \bar{A}_i на параметры в правиле, $w_{kn}, kn = \overline{1, n_{kn}}$ – веса ограничений, значения которых известны (определены);

- $C = R * T * CF \in [0, 1]$ – комплексная оценка достоверности правила, где $CF \in [0, 1]$ – достоверность правила.

Алгоритм логического вывода состоит из следующих шагов.

1. Ввод значений x_i^* входных параметров.
2. Для каждого правила $Rule_r, r = \overline{1, N}$ расчет степени срабатывания его условной части R_r .
3. Формирование конфликтного множества, включающего правила с ненулевой степенью срабатывания: $S_{conf} = \{Rule_r | R_r \neq 0\}$.
4. Для всех правил из конфликтного множества $Rule_r \in S_{conf}$ расчет совокупного веса условной части правила T_r , а также комплексной оценки C_r .
5. Разрешение конфликта – выбор правила с максимальной комплексной оценкой достоверности решения правила: $Rule_r^* : \max_{r: Rule_r \in S_{conf}} C_r$.
6. Получение значения d_j выходного параметра y правила в качестве искомого решения.

Для формирования базы знаний разработана специальная нечеткая нейронная сеть [5]. При ее обучении на экспериментальных данных формируется нейронечеткая модель, аппроксимирующая сложный объект системой нечетко-продукционных правил вида (1). На рисунке 2 приведен пример структуры нечеткой нейронной сети.

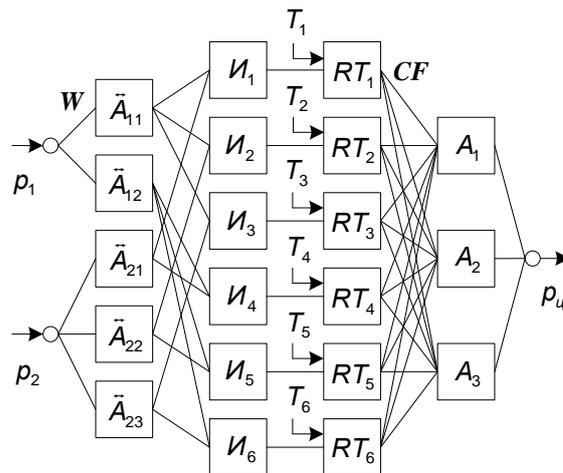


Рис. 2 - Пример структуры нечеткой нейронной сети

Структура сети определяется числом входных параметров объекта, соответствующих количеству входных параметров в обучающей выборке данных, числом значений (градаций) входных параметров, числом значений выходного параметра объекта (классов решений), а также алгоритмом логического вывода на системе правил вида (1). Первые три параметра задают количество нейронов в соответствующих слоях нечеткой нейронной сети. Последний параметр определяет число слоев сети и их функциональность.

Обучение сети основано на использовании метода градиентного спуска и позволяет за конечное число шагов произвести настройку параметров функций принадлежности в правилах, весов условий правил и их достоверности. Таким образом, нейронечеткая модель представляет собой инструмент для

формирования базы знаний в виде системы нечетко-продукционных правил.

Для оценки аппроксимирующей способности нейронечеткой модели проведено обучение нечеткой нейронной сети на данных, соответствующих задаче классификации из общедоступного источника UCI Machine Learning Repository [16]. В таблице 1 представлены характеристики наборов данных.

Таблица 1 - Характеристика исходных данных

Набор исходных данных	Число входных параметров	Типы входных параметров	Объем выборки	Число классов
BUPA Liver Disorders	6	числовые	345	2
Australian Credit Approval	14	числовые, категориальные	690	2
German Credit Data	20	числовые, категориальные	1000	2

В первом наборе представлены данные, соответствующие задаче диагностирования заболевания печени человека по результатам анализа крови. Второй набор соответствует задаче выявления подозрительных транзакций с банковскими картами. Третий набор соответствует задаче принятия решения о выдаче потребительского кредита на основании анкетных данных заемщика.

В таблице 2 приведено сравнение точности нейронечеткой аппроксимации анализируемых данных с точностью других методов классификации [1].

Таблица 2 - Сравнение методов классификации

Наборы данных	BUPA Liver Disorders	Australian Credit Approval	German Credit Data
Метод классификации			
Коллективный классификатор на нечеткой логике	0,757	0,921	0,821
Классификатор на нечеткой логике	0,725	0,891	0,794
Байесовский классификатор	0,629	0,847	0,679
Многослойный перцептрон	0,693	0,833	0,716
Бустинг	0,656	0,76	0,7
Бэггинг	0,63	0,847	0,684
Метод случайных подпространств	0,632	0,852	0,677
Коллектив нейронных сетей: простое усреднение	0,74	0,892	0,805
Коллектив нейронных сетей: равноправное голосование	0,783	0,918	0,815
Коллектив нейронных сетей: правило Борда	0,772	0,905	0,831

Коллектив нейронных сетей: многоярусное обобщение	0,785	0,925	0,852
Коллектив нейронных сетей: трехступенчатые эволюционные метод	0,804	0,947	0,857
Козволюционный метод обучения алгоритмических композиций	0,644	0,866	0,746
Нейронечеткая аппроксимация	0,762	0,859	0,827

Аппроксимирующая способность нейронечеткой модели сравнима, а иногда и превосходит точность других методов классификации, что доказывает значимость полученных результатов. Кроме того, возможность работы на некомплектных выборках данных, состоящих из значений разнотипных входных и выходных параметров, повышает эффективность предложенного подхода для аппроксимации сложных объектов с дискретным выходом.

С целью апробации произведено формирование баз знаний экспертных диагностических систем в различных социальных и технических областях человеческой деятельности. В медицине для формирования базы знаний использованы данные клинического, нейро-ортопедического, рентгенокомпьютерного томографического обследования 230 женщин в возрасте от 15 до 92 лет и 180 мужчин в возрасте от 16 до 81 года с различными синдромами поясничного остеохондроза [7]. Контрольную группу составили 20 женщин в возрасте от 20 до 70 лет и 20 мужчин в возрасте от 17 до 73 лет, никогда не страдавших поясничными болями.

Данные для анализа получены в результате клинического нейроортопедического обследования пациентов по методике В.П. Веселовского [2] и Я.Ю. Попелянского [12], а также использования компьютерного томографа SOMATOM AR.NP spiral фирмы SIEMENS при сканировании с шагом 3/3 мм. С помощью томографа изучены характеристики состояния структур позвоночно-двигательных сегментов, собственных мышц позвоночника и паравертебральных мышц на уровне L_{III}- S_I по методике М.А. Подольской, З.Ш. Нуриева [11].

Исходная выборка включала данные по 209 количественным и качественным признакам поясничного остеохондроза. С помощью нейронечеткой модели выделены 16 групп параметров и проанализированы возможные варианты зависимостей в них. Сформирована база знаний экспертной системы диагностики поясничного остеохондроза из 680 значимых нечетко-продукционных правил.

Проведем анализ и интерпретацию некоторых полученных результатов на примере сформированных правил в группе параметров диагностики клинического синдрома при поясничном остеохондрозе (см. табл. 3).

На исходных данных, соответствующих указанной группе параметров, произведено обучение нечеткой нейронной сети и сформирована система из 62 нечетко-продукционных правил, значимо

определяющих клинический синдром. В таблице 4 приведены основные результаты обучения.

Таблица 3 - Параметры диагностики синдромов поясничного остеохондроза

№	Наименование параметра	Тип параметра
1	Возраст пациента	количественный
2	Стаж статических физических нагрузок	количественный
3	Стаж динамических физических нагрузок	количественный
4	Длительность поясничной боли первого обострения	количественный
5	Длительность поясничной боли настоящего обострения	количественный
6	Скорость развития настоящего обострения	качественный
7	Выраженность поясничной боли настоящего обострения	качественный
8	Частота обострений	качественный
9	Высота межпозвонковых дисков на уровне L _{III} - S _I	количественный
10	Трабекулярное содержание кальция в позвонках L _{III} - S _I	количественный
11	Глубина протрузии межпозвонкового диска	количественный
12	Глубина грыжи межпозвонкового диска	количественный

Таблица 4 - Результаты обучения модели

Параметры	Значения	
	женщины	мужчины
Количество сформированных правил	31	31
Объем обучающей выборки – число пациентов	230	180
Точность диагностики на обучающей выборке, %	87,83	89,44
Объем тестовой выборки – число пациентов контроля	20	20
Точность диагностики на тестовой выборке, %	95	90

Точность диагностики на тестовой выборке данных составила 90%, что указывает на высокую согласованность диагнозов опытного врача-эксперта и результатов диагностики на системе сформированных нечетко-продукционных правил.

Результаты исследования используются в научно-исследовательской работе кафедры реабилитации и спортивной медицины Казанской государственной медицинской академии под руководством кандидата медицинских наук, доцента Подольской М.А. для сбора, систематизации и анализа медицинских данных по теме «Патогенез, диагностика и восстановительное лечение периферических заболеваний нервной системы». Сформированная база знаний используется в разрабатываемой экс-

пертной системе диагностики, прогнозирования и лечения поясничного остеохондроза.

В нефтяной отрасли сформирована база знаний и разработана экспертная система для оперативного автоматического обнаружения утечек из водоводов (порывов) в системах поддержания пластового давления [3]. Для формирования базы знаний использовалась статистическая информация по водоводам цеха ППД-2 НГДУ «ДжалильНефть» ОАО «ТатНефть».

В качестве исходных данных для анализа выступали значения следующих параметров:

а) получасовые расходы жидкости на водоводах (м³):

- Q_0 – текущий расход по водоводу;

- Q_1 – расход по водоводу за предыдущие

пол часа;

- Q_2 – расход по водоводу за предыдущий

час;

б) показатели давления (МПа):

- P_0 – текущее давление в коллекторе;

- P_1 – предыдущее давление в коллекторе.

Для подготовки обучающих выборок рассчитаны относительные отклонения расхода жидкости по каждому из водоводов за пол часа и за час:

$$\varepsilon = \frac{Q_0 - Q}{Q} * 100\%,$$

где $Q \in \{Q_1, Q_2\}$.

Кроме того, рассчитаны изменения давлений в коллекторе: $\Delta P = P_1 - P_0$.

Обучающие выборки включали значения следующих входных параметров:

- ε_1 – отклонение расхода жидкости по водоводу за 0,5 часа;

- ε_2 – отклонение расхода жидкости по водоводу за 1 час;

- ΔP – уменьшение давления в коллекторе.

В качестве целевого выступал параметр D (состояние водовода), принимающий одно из двух возможных значений: «норма» или «порыв». Указанные параметры составляли группы «входы - выход» по каждому водоводу кустовых насосных станций, общее число которых составило 28, число водоводов – 303, суммарный объем скважин – 656. В результате обучения нечетких нейронных сетей сформировано 28 систем нечетко-продукционных правил, общее число которых составило 342.

Разработанная экспертная диагностическая система успешно прошла тестирование на лабораторных стендах «ТатАСУНефть», этап опытной эксплуатации и внедрена в промышленную эксплуатацию на диспетчерском пункте ЦППД-2 НГДУ «ДжалильНефть» ОАО «ТатНефть». При работе система показала высокую эффективность используемого алгоритма оперативного обнаружения порывов, что подтверждено 100%-м уровнем выявления реальных аварийных ситуаций на водоводах кустовых насосных станций. Использование системы позволило не только оперативно в автоматизированном режиме выявлять утечки на водоводах, но

и снизить смысловую нагрузку на специалистов-технологов и диспетчеров при анализе информации по выявленным авариям. В результате внедрения экспертной системы в эксплуатацию повысилась эффективность диагностики состояния водоводов и, как следствие, улучшились технико-экономические показатели по закачке жидкости в нефтяной пласт.

В области информационной безопасности сформирована база знаний для фильтрации электронных почтовых сообщений [4]. Данная задача сводится к классификации писем на категории «спам» и «не спам» и требует анализа значений следующих параметров: длина темы сообщения, количество составных частей письма, количество строк письма, количество цифр, количество ссылок, количество пустых строк и количество цветов. Сформированная база знаний из 32 нечетко-продукционных правил успешно прошла апробацию на рабочих станциях информационных систем Управления связи, специальной техники и автоматизации Министерства внутренних дел по Республике Татарстан и используется в составе диагностической экспертной системы предварительного выявления электронных писем несанкционированной массовой рассылки.

Использование системы позволило повысить уровень защищенности корпоративной сети и ее информационных ресурсов, сократить затраты на оплату трафика входящей электронной почты, сократить время и снизить нагрузку на специалистов при анализе входящей корреспонденции и, как следствие, повысить их работоспособность.

Описанный в данной работе подход является примером реализации информационного подхода к моделированию. Предложенная нейронечеткая модель, обучаясь на экспериментальных данных, формирует систему нечетко-продукционных правил для аппроксимации сложных объектов с дискретным выходом. Таким образом, модель является эффективным средством для формирования баз знаний и может успешно применяться в составе экспертных диагностических систем.

Литература

1. Бухтояров В.В. Трехступенчатый эволюционный метод формирования коллективов нейронных сетей для решения задач классификации // Программные продукты и системы. – 2012. – № 4. – С. 101-106.

2. Веселовский В.П. Практическая вертеброневрология и мануальная терапия. Рига. 1991. 344 с.
3. Глова В.И., Аникин И.В., Катасёв А.С. Система предупреждения аварий оборудования в процессах поддержания пластового давления // Вестник Казанского государственного технического университета им. А.Н. Туполева, № 2, 2006. – С. 46-49.
4. Катасёв А.С. Формирование базы знаний системы фильтрации электронных почтовых сообщений // Научно-технический вестник Поволжья, № 5, 2013. – С. 191-194.
5. Катасёв А.С., Ахатова Ч.Ф. Нейронечеткая система обнаружения продукционных зависимостей в базах данных // Программные продукты и системы. – 2011. – № 3. – С. 26-32.
6. Катасёв А.С., Газимова Д.Р. Инвариантная нечетко-продукционная модель представления знаний в экспертных системах // Вестник КГТУ им. А.Н. Туполева. – 2011. – № 1. – С. 142-148.
7. Катасёв А.С., Подольская М.А. Применение систем искусственного интеллекта для диагностического процесса в вертеброневрологии // Казанский медицинский журнал, том 88, № 4, 2007. – С. 346-351.
8. Кирпичников А.П., Осипова А.Л., Ризаев И.С. Повышение аналитических возможностей баз данных // Вестник Казан. технол. ун-та. – 2012. – № 3. – С. 157-160.
9. Паклин Н.Б. Адаптивные модели нечеткого вывода для идентификации нелинейных зависимостей в сложных системах: Дисс. на соиск. уч. степ. к-та техн. наук. – Ижевск, 2004. – 162 с.
10. Паклин Н.Б., Орешков В.И. Бизнес-аналитика: от данных к знаниям: учебное пособие. – 2-е изд., испр. – СПб.: Питер, 2013. – 704 с.: ил.
11. Подольская М.А., Нуриев З.Ш. Компьютерно-томографическое исследование паравертебральных мышц на поясничном уровне при дистрофических вертеброгенных заболеваниях // Медицинская визуализация. 2004. № 4. С. 127-136.
12. Попелянский Я.Ю. Вертебральные синдромы поясничного остеохондроза. – Казань, 1974. – 284 с.
13. Ротштейн А.П. Интеллектуальные технологии идентификации: нечеткая логика, генетические алгоритмы, нейронные сети. – Винница: УНИВЕРСУМ - Винница, 1999. – 320 с.
14. Титов А.Н., Нуриев Н.К., Тазиева Р.Ф. Оценка параметров вероятностной модели по экспериментальным данным // Вестник Казан. технол. ун-та. – 2013. – № 19. – С. 324-330.
15. Штовба С.Д. Классификация объектов на основе нечеткого логического вывода // Exponenta Pro - Математика в приложениях. – 2004. – № 1(5). – С. 68-69.
16. Bache K., Lichman M. UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science. 2013.

© Л. Ю. Емалетдинова – д-р техн. наук, проф. каф. прикладной математики и информатики КНИТУ-КАИ, lilia@stcline.ru; А. С. Катасёв – канд. техн. наук, доц. каф. систем информационной безопасности КНИТУ-КАИ, kat_726@mail.ru; А. П. Кирпичников – д-р физ.-мат. наук, зав. каф. интеллектуальных систем и управления информационными ресурсами КНИТУ, kirpichnikov@kstu.ru.