

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ, ЧИСЛЕННЫЕ МЕТОДЫ И КОМПЛЕКСЫ ПРОГРАММ

УДК 004.89

НЕЙРОНЕЧЕТКАЯ МОДЕЛЬ И ПРОГРАММНЫЙ КОМПЛЕКС ФОРМИРОВАНИЯ БАЗ ЗНАНИЙ ДЛЯ ОЦЕНКИ СОСТОЯНИЯ ОБЪЕКТОВ

Статья поступила в редакцию 18.02.2022, в окончательном варианте – 24.02.2022.

Катасёва Дина Владимировна¹, Казанский национальный исследовательский технический университет им. А.Н. Туполева-КАИ, 420111, Российская Федерация, г. Казань, ул. К. Маркса, 10, аспирант, ORCID: 0000-0001-6141-8329, e-mail: DVKataseva@kai.ru

Описана задача оценки состояния объектов. Для повышения эффективности ее решения и снижения влияния человеческого фактора актуализирована целесообразность использования систем поддержки принятия решений и формирования баз знаний нечетко-продукционного типа. В качестве инструмента их формирования предложено использовать нейронечеткую модель на основе обучения нечеткой нейронной сети. Рассмотрены особенности исходных данных для обучения. Предложен вид нечетко-продукционных правил, составляющих базу знаний. Рассмотрены этапы разработки нейронечеткой модели, включающие инициализацию и настройку значений параметров нечеткой нейронной сети. Описан разработанный метод ее обучения, основанный на использовании генетического алгоритма. Для его реализации рассмотрены вопросы выбора и кодирования параметров обучения на примере треугольных функций принадлежности. Предложена фитнес-функция и критерий выбора лучшей хромосомы в генетическом алгоритме. На основе предложенного математического обеспечения разработан программный комплекс для формирования баз знаний. Представлена его структура, состав программных модулей и их функциональность. Приведены результаты апробации программного комплекса на примере формирования базы знаний для подбора геолого-технических мероприятий на нефтяном месторождении. Результаты апробации подтвердили эффективность предложенного подхода и возможность его использования при формировании баз знаний и построении систем поддержки принятия решений для оценки состояния объектов в различных предметных областях.

Ключевые слова: оценка состояния объекта, база знаний, нечеткая нейронная сеть, нейронечеткая модель, нечетко-продукционное правило, функция принадлежности, генетический алгоритм, система поддержки принятия решений, геолого-технические мероприятия, нефтяное месторождение

NEURO-FUZZY MODEL AND SOFTWARE COMPLEX FOR FORMING KNOWLEDGE BASES FOR OBJECTS STATE ASSESSING

The article was received by the editorial board on 18.02.2022, in the final version – 24.02.2022.

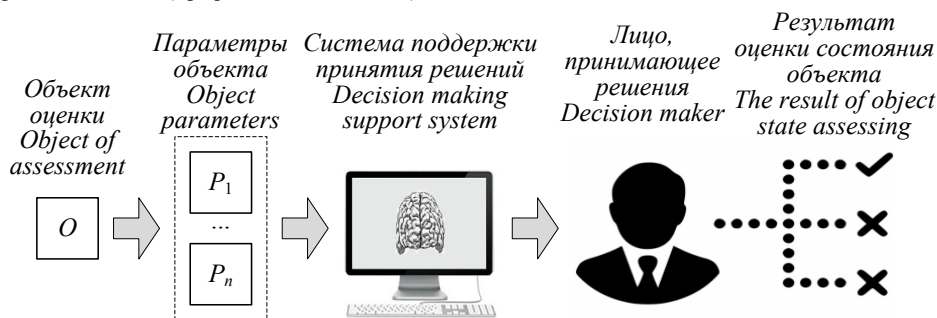
Kataseva Dina V., Kazan National Research Technical University named after A.N. Tupolev-KAI, 10 K. Marx St., Kazan, 420111, Russian Federation, postgraduate student, ORCID: 0000-0001-6141-8329, e-mail: DVKataseva@kai.ru

The task of objects state assessing is described. To improve the efficiency of its solution and reduce the human factor influence, the expediency of using decision-making support systems and the formation of fuzzy-production type knowledge bases has been updated. As a tool for their formation, it is proposed to use a neuro-fuzzy model based on a fuzzy neural network training. The features of the initial data for training are considered. A type of fuzzy-production rules that make up the knowledge base is proposed. The stages of neuro-fuzzy model construction, including the initialization and setting of the values of fuzzy neural network parameters, are considered. The developed method of its training based on the use of a genetic algorithm is described. For its implementation, the issues of choosing and coding learning parameters are considered on the example of triangular membership functions. A fitness function and a criterion for choosing the best chromosome in the genetic algorithm are proposed. Based on the proposed methods, a software package for the knowledge bases formation has been developed. Its structure, composition of program modules and their functionality are presented. The results of the software package approbation are presented on the example of the knowledge base formation for the selection of geological and technical measures in oil fields. The approbation results confirmed the effectiveness of the proposed approach and the possibility of its use in the knowledge bases formation and the decision-making support systems construction for objects state assessing.

Keywords: object state assessment, knowledge base, fuzzy neural network, neuro-fuzzy model, fuzzy production rule, membership function, genetic algorithm, decision support system, geological and technical measures, oil field

¹ Научный руководитель: Исмагилов Ильяс Идрисович, Казанский национальный исследовательский технический университет им. А.Н. Туполева-КАИ, 420111, Российская Федерация, г. Казань, ул. К. Маркса, 10, доктор технических наук, профессор, ORCID: 0000-0002-0446-8204, e-mail: iisimag@mail.ru

Graphical annotation (Графическая аннотация)



Введение. В настоящее время в различных сферах человеческой деятельности возникает необходимость решения задач, связанных с оценкой состояния объектов [8]. Как правило, при решении таких задач ключевую роль играет человек – лицо, принимающее решения (ЛПР) [3]. На основе анализа значений параметров объекта P_1, \dots, P_n он формирует результат оценки его состояния, выбирая оптимальное решение из множества допустимых. При этом ЛПР опирается на свое мнение, подкрепленное знаниями, опытом и интуицией.

Однако такая схема оценки сопряжена с рядом трудностей в силу наличия человеческого фактора. В качестве ЛПР может выступать как квалифицированный, так и неквалифицированный пользователь, ЛПР может ошибаться из-за невнимательности, усталости и т.д. Все это может приводить к ошибкам. Поэтому в настоящее время для повышения эффективности этого процесса актуально использование систем поддержки принятия решений (СППР) по оценке состояния объектов [1]. Рекомендации, поступающие от СППР, помогают ЛПР принять правильное решение, не ошибиться в выборе текущего состояния объекта.

Основными преимуществами применения СППР в задаче оценки состояния объекта являются [11]:

- 1) высокая скорость обработки информации при принятии решений;
- 2) адекватность рекомендуемых решений исходным данным, характеризующим оцениваемый объект;
- 3) снижение трудоемкости процесса принятия решений для человека.

Высокая скорость обработки информации достигается за счет автоматизации вычислений при анализе значений параметров объекта, выполнении процедуры логического вывода, сопоставлении входных данных с имеющимися в базе знаний системы правилами принятия решений. Адекватность рекомендуемых решений исходным данным об объекте оценки обусловлена тем, что интеллектуальная система всегда работает по одному алгоритму логического вывода и применяет для этого заранее сформированную базу знаний, адекватность которой обеспечивается в процессе ее формирования. Снижение трудоемкости выражается в уменьшении интеллектуальной нагрузки на ЛПР в процессе принятия решений.

Применение СППР главным образом приводит к снижению влияния человеческого фактора на процесс принятия решений и позволяет автоматизировать данный процесс.

При этом к СППР предъявляются следующие основные требования:

- 1) высокая точность оценки состояния объектов;
- 2) умение обосновать рекомендуемые решения в виде, пригодном для интерпретации ЛПР.

Высокая точность оценки состояния объектов должна достигаться за счет формирования адекватных баз знаний. Умение обосновывать рекомендуемые решения необходимо для их понимания и интерпретации человеком. Это, в свою очередь, повышает уровень доверия ЛПР к предлагаемым системой решениям, что позволяет сделать обоснованный выбор окончательного решения.

Одним из наиболее эффективных инструментов формирования баз знаний являются нечеткие нейронные сети (ННС) [2]. Рассмотрим особенности их использования.

Использование нечетких нейронных сетей для формирования баз знаний. Особенностью всех ННС является их гибридный характер [5]. Подобно классическим нейронным сетям они обучаются на исходных данных, формируя базу знаний нечетко-продукционного типа. Правила базы знаний могут быть использованы для решения поставленных задач, в данном случае – для задачи оценки состояния объектов.

На рисунке 1 представлена схема использования ННС для формирования базы знаний.

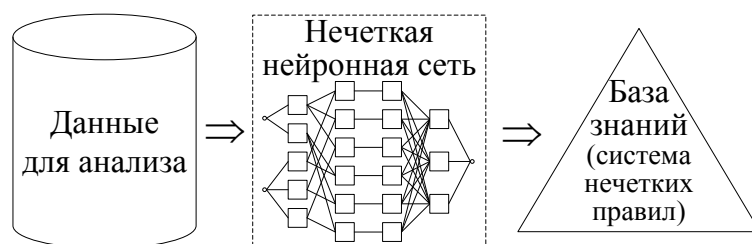


Рисунок 1 – Схема использования нечеткой нейронной сети для формирования базы знаний

Для реализации представленной схемы необходимо иметь набор данных, характеризующих решаемую задачу. Данные для анализа могут быть представлены в виде, как показано в таблице 1.

Таблица 1 – Общий вид таблицы данных для анализа

Входные переменные					Выходная переменная
1	2	3	...	n	

Таблица данных объемом N содержит n входных и одну выходную переменные, характеризующие оцениваемый объект. Каждая строка такой таблицы соответствует определенному объекту и содержит значения входных и выходной переменных. Значением выходной переменной является конкретное состояние объекта, в котором он находился при указанных условиях (конкретных значениях входных переменных). Следовательно, имея характеристики N объектов, можно сформировать исходную таблицу данных для анализа. При этом ННС выступает в качестве инструмента анализа данных [4].

Таким образом, для реализации описанной схемы формирования базы знаний на основе нейронечеткого подхода необходимо решить следующие основные задачи:

- 1) определить особенности исходных данных, характеризующих оцениваемые объекты и используемых для принятия решений;
- 2) предложить вид нечетких правил, учитывающих особенности исходных данных;
- 3) разработать нейронечеткую модель для формирования правил выбранного вида.

Особенности исходных данных. При принятии решений используются входные данные, характеризующие оцениваемые объекты. Объекты оценки могут быть различной природы. При этом данные, описывающие оцениваемые объекты, могут быть получены из различных источников. Поэтому для эффективного решения поставленной задачи необходимо понимать и учитывать их особенности, а именно возможную их разнотипность и некомплектность.

Разнотипность данных означает, что часть входных переменных в таблице для анализа является количественной и принимает непрерывные или дискретные значения, лежащие на числовой оси. Остальные входные переменные являются качественными и принимают категориальные (номинальные, порядковые или бинарные) значения. Некомплектность данных, используемых для принятия решений, означает ситуацию, когда в процессе оценки состояния объекта на входе СППР появляются данные с пропущенными значениями.

Вид нечетких правил. Для учета особенностей входных данных, используемых при оценке состояния объектов, предложен следующий вид нечетко-продукционных правил:

$$\text{If } x_1 = \tilde{A}_1 \& x_2 = \tilde{A}_2 \& \dots x_i = \tilde{A}_i \& \dots x_n = \tilde{A}_n \text{ Then } y = B[CF], \quad (1)$$

где x_i ($i=1..n$) – входные переменные; \tilde{A}_i – четкие ($\tilde{A}_i = A_i$) и нечеткие ($\tilde{A}_i = \tilde{A}_i$) ограничения на значения входных переменных, позволяющие обрабатывать, соответственно, качественные и количественные входные данные; A_i – четкие значения входных переменных; $\tilde{A}_i = \{x_i, \mu_{\tilde{A}_i}(x_i)\}$ – нечеткие значения входных переменных; y – выходная переменная; B – состояние объекта; $CF \in [0; 1]$ – вес правила.

Достоинства нечетко-продукционных правил вида (1) заключаются в следующем:

- 1) позволяют обрабатывать разнотипные входные данные;
- 2) поддаются однозначной лингвистической интерпретации;
- 3) характеризуются весом CF (для оценки состояния объекта).

Рассмотрим предложенную в работе нейронечеткую модель.

Разработка нейронечеткой модели для формирования баз знаний. Для построения любой интеллектуальной модели на основе методов машинного обучения требуется выполнить ряд последовательных этапов. Рассмотрим этапы построения нейронечеткой модели (рис. 2).

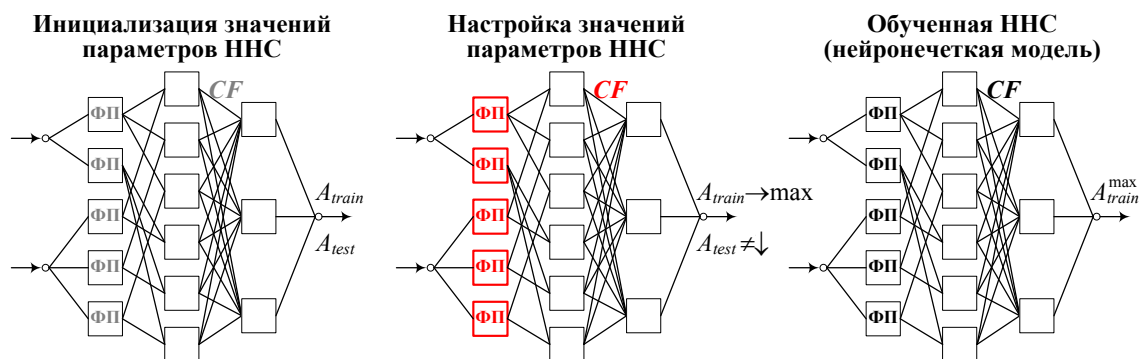


Рисунок 2 – Этапы построения нейронечеткой модели

Для построения нейронечеткой модели требуется выполнить этапы инициализации и настройки значений параметров ННС. На первом этапе необходимо определить начальные значения параметров функций принадлежности (ФП) и начальные значения весов CF . Также требуется рассчитать начальные значения точности классификации ННС на обучающей (A_{train}) и тестовой (A_{test}) выборках данных.

На этапе настройки производится так называемый «тюнинг» модели [6], в результате которого определяются оптимальные значения указанных параметров с учетом следующего критерия оптимизации:

$$A_{train} = \frac{N_{train_t}}{N_{train}} \rightarrow \max \mid A_{test} = \frac{N_{test_t}}{N_{test}} \neq \downarrow, \quad (2)$$

где N_{train} – объем обучающей выборки; N_{train_t} – число правильно классифицированных обучающих данных ($N_{train_t} \leq N_{train}$); N_{test} – объем тестовой выборки; N_{test_t} – число правильно классифицированных тестовых данных ($N_{test_t} \leq N_{test}$); « $\neq \downarrow$ » – условие неуменьшения точности классификации ННС на тестовой выборке данных.

Следовательно, при реализации этого этапа необходимо максимизировать точность классификации A_{train} нечеткой нейронной сети на обучающей выборке данных в условиях контроля ее переобучения (неуменьшения точности классификации ННС A_{test} на тестовой выборке).

Результатом выполнения указанных этапов является построенная нейронечеткая модель с настроенными значениями параметров. Для реализации отдельных этапов ее построения требуется разработка соответствующих методов и алгоритмов.

Инициализация является первичным этапом построения нейронечеткой модели. Рассмотрим его реализацию на примере треугольных ФП, используемых для формализации нечетких ограничений на значения входных количественных переменных в нечетких правилах. Выбор треугольных ФП обусловлен широтой и эффективностью их практического использования [10, 13], а также простотой задания и лингвистической интерпретацией человеком, что требуется для принятия обоснованных решений.

На рисунке 3 показаны примеры определения начальных значений параметров ФП.

При любом количестве выбираемых нечетких градаций используется равномерный метод гранулирования. Такой подход является классическим и позволяет максимально просто определить начальные значения параметров каждой функции принадлежности [15].

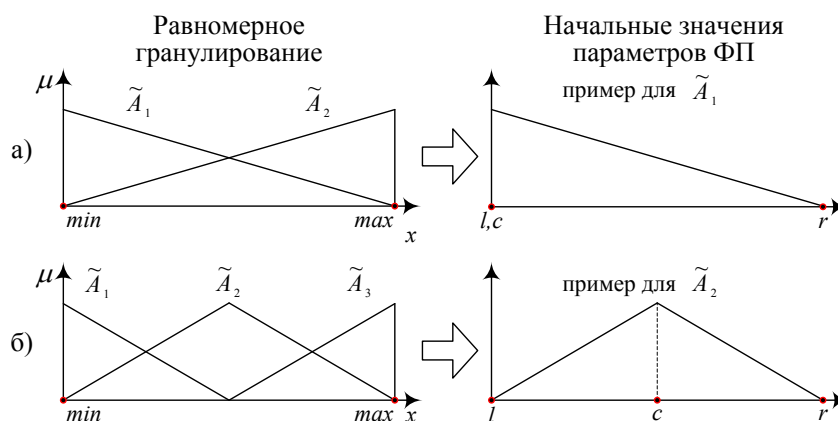


Рисунок 3 – Определение начальных значений параметров функций принадлежности

Так, например, в случае 2-х нечетких градаций задаются две треугольные ФП на отрезке $[min; max]$, как показано на рисунке 3а. Границы этого отрезка соответствуют минимальному и максимальному значениям входного параметра x в обучающей выборке данных. Это позволяет определить начальные значения параметров ФП: l – левое основание, c – моду и r – правое основание. На рисунке 3а справа показан пример определения начальных значений параметров для первой функции принадлежности. Аналогично определяются начальные значения параметров ФП для 3-х (рис. 3б) и большего числа градаций.

После выполнения этого этапа необходимо обучить нечеткую нейронную сеть путем настройки значений ее параметров. Рассмотрим разработанные метод и алгоритмы, используемые для обучения ННС.

Существует большое количество методов оптимизации, используемых в машинном обучении при построении интеллектуальных моделей. Так, например, при обучении нейронных и нечетких нейронных сетей большое распространение получили градиентные методы [14], используемые в алгоритме обратного распространения ошибки, который стал классическим для построения таких моделей. Данный алгоритм предполагает вычисление и минимизацию среднеквадратичной ошибки выхода сети на основе вычисления градиентов по настраиваемым параметрам модели.

Однако в предложенной нечеткой нейронной сети функция ошибки принимает дискретные значения и зависит от дискретных параметров, поэтому применение понятия «градиент» для нее будет некорректным. Кроме того, градиент вычисляется для гладких функций. Используемые в работе треугольные функции принадлежности кусочно-линейны, т. е. не являются гладкими на всей области определения. Поэтому для обучения ННС предложено использовать генетический алгоритм (ГА) [9]. Такие алгоритмы получили широкое распространение. Они не требуют использования гладких ФП и могут быть реализованы с использованием различных эвристических приемов, связанных с кодированием параметров задачи, выполнением генетических операторов и выбором критериев для нахождения квазиоптимальных решений поставленных задач.

На рисунке 4 представлена схема разработанного метода обучения ННС с использованием генетического алгоритма.

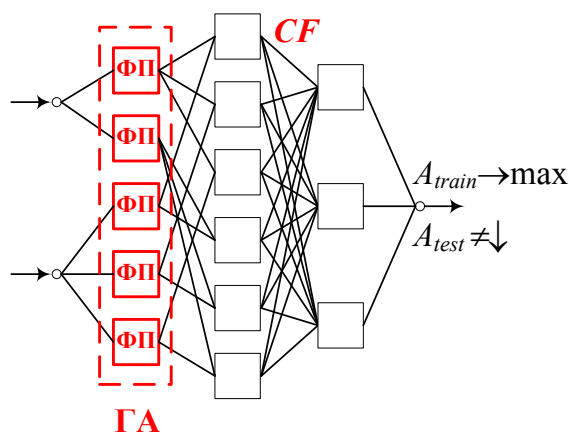


Рисунок 4 – Схема метода обучения нечеткой нейронной сети

Из рисунка видно, что ГА используется для настройки значений параметров ФП в ННС. При этом на каждой итерации генетической оптимизации происходит уточнение значений весовых коэффициентов CF . Целью обучения является максимизация A_{train} (точности классификации обучающих данных) при условии неуменьшения A_{test} (точности классификации тестовых данных).

Таким образом, для обучения ННС разработан метод итерационной генетической оптимизации, позволяющий настраивать значения параметров ФП и по результатам настройки определять итоговые значения весовых коэффициентов CF .

Рассмотрим вопросы выбора и кодирования параметров ФП, представленных на рисунке 5.

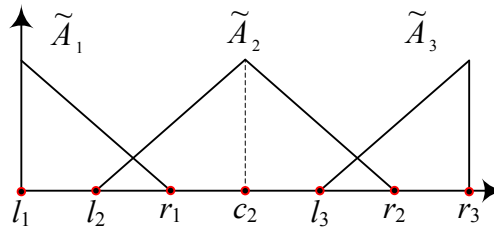


Рисунок 5 – Функции принадлежности и их параметры

На рисунке представлены три треугольные функции принадлежности:

- \tilde{A}_1 с параметрами (l_1, r_1) ;
- \tilde{A}_2 с параметрами (l_2, c_2, r_2) ;
- \tilde{A}_3 с параметрами (l_3, r_3) .

Следует отметить, что для задания крайних функций принадлежности (в данном случае это \tilde{A}_1 и \tilde{A}_3) достаточно иметь два параметра (в данном случае это l – левое основание и r – правое основание). Во всех остальных случаях ФП задаются тремя параметрами: l – левое основание, c – мода и r – правое основание.

Очевидно, что не все из представленных параметров $\{l_1, l_2, r_1, c_2, l_3, r_2, r_3\}$ являются настраиваемыми. Следовательно, необходимо выбирать такие параметры p , значения которых можно настраивать. Рассмотрим вопрос выбора настраиваемых параметров.

Обучение ННС заключается в настройке значений параметров ее функций принадлежности. При этом значения крайних параметров (в данном случае это параметры l_1 и r_3) не могут изменяться, иначе нарушится интерпретируемость соответствующих крайних функций принадлежности. Следовательно, $p \in \{l_2, r_1, c_2, l_3, r_2\}$. Значения именно этих параметров могут и должны изменяться.

Следующим вопросом является определение интервалов значений для настраиваемых параметров. Поскольку функции принадлежности должны быть интерпретируемы, упорядочены и не иметь точек разрыва, то в качестве указанных интервалов параметров $p \in (\min; \max)$ целесообразно выбрать следующие:

$$l_2 \in (l_1; r_1), r_1 \in (l_2; l_3), c_2 \in (l_2; r_2), l_3 \in (r_1; r_2), r_2 \in (l_3; r_3).$$

Далее необходимо закодировать значения настраиваемого параметра в заданном интервале. Для этого каждый интервал равномерно разобьем на $2^m + 1$ отрезков, где m – целое число. Тогда каждый параметр сможет принимать одно из 2^m значений:

$$H = \{H_1, \dots, H_z, \dots, H_Z\} = \{\overbrace{(0, 0, \dots, 0)}^m, \overbrace{(0, 0, \dots, 1)}^m, \dots, \overbrace{(1, 1, \dots, 1)}^m\}, z = 1..Z, Z = 2^m,$$

где $H_1 = (0, 0, \dots, 0)$ – хромосома, соответствующая наименьшему значению параметра; $H_Z = (1, 1, \dots, 1)$ – хромосома, соответствующая его наибольшему значению; m – длина хромосомы (количество ее ген).

На рисунке 6 показан пример кодирования (для $m = 4$).

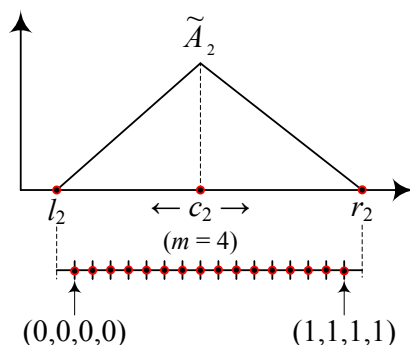


Рисунок 6 – Пример кодирования значений параметра функции принадлежности

Интервал $(l_2; r_2)$ ограничивает возможные значения для настраиваемого параметра c_2 . В данном случае этот параметр может принимать одно из 16 значений (так как $m = 4$, то $2^m = 2^4 = 16$).

Завершающим вопросом, требующим решения, является задание фитнес-функции для оценки значений настраиваемого параметра. Так как целью обучения ННС является максимизация точности ее классификации на обучающей выборке данных, то в качестве фитнес-функции целесообразно задать следующую:

$$F(H_z) = \frac{N_{train_t}}{N_{train}}, \quad (z = 1 \dots Z, Z = 2^m). \quad (3)$$

Лучшая хромосома максимизирует точность классификации ННС на обучающей выборке (в соответствии с (3)) при условии неумножения ее точности классификации A_{test} на тестовой выборке:

$$F(H_z) \rightarrow \max_{H_z} | A_{test}(H_z) \nrightarrow, \quad A_{test}(H_z) = \frac{N_{test_t}}{N_{test}}. \quad (4)$$

Таким образом, выбор настраиваемых параметров функций принадлежности, определение интервалов и кодирование их возможных значений, а также задание фитнес-функции с критерием выбора лучшей хромосомы позволили реализовать классический ГА [7] для построения нейронечеткой модели.

Программный комплекс формирования баз знаний. Описанные модель и алгоритмы реализованы в программном комплексе, создание которого осуществлялось на базе интегрированной среды разработки Visual Studio 2019 на языке C#. Программный комплекс имеет модульную структуру (рис. 7).

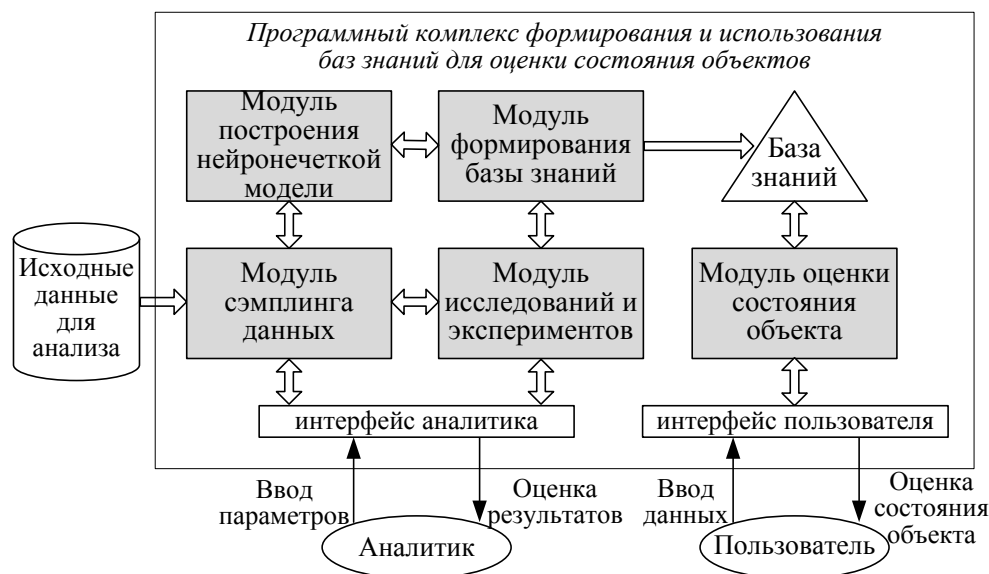


Рисунок 7 – Структура программного комплекса

Программный комплекс состоит из следующих модулей:

- 1) модуль сэмпинга данных (производит загрузку исходных данных, их визуализацию, формирование выборок данных для анализа);
- 2) модуль построения нейронечеткой модели (строит и визуализирует структуру ННС, инициализирует значения параметров ННС, производит ее обучение);
- 3) модуль формирования базы знаний (формирует систему правил на основе нейронечеткой модели, визуализирует правила базы знаний);
- 4) модуль оценки состояния объектов (производит оценку состояния объектов, визуализирует результаты оценки);
- 5) модуль исследований и экспериментов (позволяет производить исследования нейронечеткой модели и методов ее построения, визуализирует результаты исследований).

Каждый программный модуль обладает соответствующими функциональными возможностями, в совокупности составляющими возможности всего программного комплекса. Кроме того, программный комплекс имеет интерфейсы для работы с ним аналитика и конечного пользователя. Под управлением аналитика производится формирование базы знаний. Пользователь использует сформированную базу знаний непосредственно для оценки состояния объектов.

Следует отметить, что программный комплекс прошел апробацию при формировании баз знаний в различных предметных областях, в частности в нефтяной отрасли. Рассмотрим результаты его апробации.

Апробация программного комплекса в нефтяной отрасли. Одной из актуальных задач в нефтяной отрасли является подбор геолого-технических мероприятий (ввод скважин в эксплуатацию) на нефтяном месторождении. Для обучения ННС и формирования базы знаний использовались исходные данные, описываемые следующими параметрами:

- 1) общие сведения о каждой скважине:
 - порядковый номер скважины;
 - промысловый номер добывающей скважины;
 - год начала отбора и др.;
- 2) геолого-геофизические данные нефтяного пласта:
 - проницаемость абсолютная;
 - пористость;
 - нефтенасыщенность начальная и др.;
- 3) параметры на год начала отбора:
 - площадь зоны дренирования;
 - начальные балансовые запасы нефти;
 - текущие балансовые запасы нефти и др.

Исходные данные были представлены по 46 объектам нефтяного месторождения с карбонатными коллекторами. Объем данных составил 3398 записей.

В таблице 2 представлен фрагмент исходных данных для анализа.

В данной таблице в качестве входных параметров указаны следующие:

- pronic – абсолютная проницаемость породы;
- poris – пористость породы;
- nasich – начальная нефтенасыщенность;
- tolsh – начальная нефтенасыщенная толщина;
- plosk – площадь зоны дренирования;
- bal1 – начальные балансовые запасы нефти;
- bal2 – текущие балансовые запасы нефти;
- kohvat – коэффициент охвата заводнением;
- rnagn – расстояние до ближайшей нагнетательной скважины;
- debsos – средний дебит соседних скважин;
- obvod – минимальная обводненность соседних скважин.

Единственный выходной параметр «result» характеризует описываемую скважину как эффективную в применении (result = 1 при дебите скважины > 5) или неэффективную (result = 0 при дебите скважины ≤ 5).

Таблица 2 – Фрагмент исходных данных по скважинам

pronc	poris	nasich	tolsh	plosh	bal1	bal2	kohvat	rnagn	debsos	obvod	result
56,7	0,131	0,685	5,8	12	22	20,7	0,1375	7586	0,2	0,5	0
12	0,119	0,656	4,4	12	14,6	11,8	0,3531	17111	0,1	0,597	0
56,7	0,131	0,685	1,6	0,1	0	0	0,3543	0	0	0	0
56,7	0,131	0,685	6,4	12	24,3	23,8	0,0418	12390	0,4	0,794	0
56,7	0,131	0,685	8,4	4,4	5,8	5,8	0,005	5798	1,4	0,095	1
87	0,162	0,695	2,8	12	13,3	11,6	0,2483	12366	0	0	0
56,7	0,131	0,685	4,2	12	15,9	15	0,1424	11272	0	0	0
54,8	0,133	0,683	7,1	12	21	19,8	0,13	7855	0,5	0,955	0
30	0,148	0,49	1,4	12	4,3	4	0,2084	1005	0	0	0
10,4	0,108	0,644	5,2	12	11,3	10,5	0,3722	13787	0,3	0,773	0
50,7	0,138	0,714	9,9	12	34,4	31,1	0,0925	4270	0	0	0
56,7	0,131	0,685	4,6	7,3	5,3	4,9	0,1638	3190	0,9	0,837	0
45,3	0,12	0,681	14,9	12	32,3	31,4	0,0719	34181	1	0,265	0
2	0,097	0,796	16,1	12,4	34,8	29,4	0,4396	25489	0	0,065	0
3	0,116	0,749	16,9	12	60,4	59,9	0,0136	4034	1,1	0,709	0
13,2	0,121	0,687	3,7	12	14,8	14,3	0,0616	0	0	0	0
1	0,096	0,565	2,8	11,3	8,1	7,6	0,1563	0	0,3	0,755	0
6	0,127	0,66	19,7	0,4	1,4	1,4	0,0284	0	9,5	0,102	0
5,8	0,111	0,792	6,5	12	24,5	24,3	0,0097	0	1,4	0,215	1
4	0,123	0,6	5,8	12	24,3	23,7	0,0469	0	0	0	0
13,9	0,13	0,77	4,7	12	26,7	25,8	0,0529	0	0,9	0,794	1
13,9	0,094	0,74	5,2	13,1	28,5	26,1	0,14	0	0,2	0,778	0
1,3	0,106	0,623	5,8	12	16	14,9	0,243	0	0,2	0,858	0

Перед формированием базы знаний произведена подготовка исходных данных к анализу. Никакой метод машинного обучения не даст положительного результата, если данные будут плохого качества. Поэтому важным является оценка качества исходных данных и их подготовка к анализу.

В наборе исходных данных встречались выбросы и аномалии. При подготовке исходных данных к анализу выполнен их поиск и корректировка, поскольку они являются одним из факторов, снижающих качество данных и достоверность результатов анализа. Алгоритм поиска выбросов и аномалий позволил оценить качество исходных данных и выработать рекомендации по его улучшению. Всего в исходных данных объемом 3398 строк было найдено 917 строк с выбросами и 211 строк с аномалиями. Параметр «tir» оказался непригодным для анализа. Единственный параметр «obvod» оказался полностью пригодным к анализу. К остальным значениям параметров были применены методы повышения качества исходных данных [12]: удаления записей с аномальными значениями и заменой выбросов на медианное значение. После предобработки объем исходных данных сократился на 211 строк (с 3398 до 3187) в соответствии с количеством удаленных записей с аномальными значениями.

Следующим этапом подготовки данных явился корреляционный анализ между параметрами. При построении интеллектуальной модели целесообразно оценить зависимости в исходных данных для анализа, а именно выбрать систему входных признаков, значимо влияющих на выходной результат. Данная задача актуальна при обучении ННС, способных эффективно обучаться в пространстве небольшой (до 10–12 входных параметров) размерности.

В результате корреляционного анализа данных произведен выбор входных параметров. В таблице 3 представлены значения рассчитанных коэффициентов корреляции.

Из таблицы видно, что не все входные параметры имеют корреляционную зависимость с выходным. Так, например, такие параметры, как «проницаемость», «пористость», «средний дебит соседних скважин» и «минимальная обводненность соседних скважин» имеют относительно высокую корреляцию с выходным параметром «результат». Это означает, что указанные параметры целесообразно использовать в обучении. При этом порог корреляции, определяющий количество и состав входных параметров, подбирался экспериментально.

После подготовки исходных данных и их анализа произведено формирование баз знаний с различным набором комбинаций параметров. Проведены следующие исследования:

- оценка влияния порога корреляции в исходных данных на результаты моделирования;
- оценка влияния порога эффективности скважины на результаты обучения ННС;
- оценка влияния вариантов расчета выходной переменной (первый ненулевой дебит скважины за последние 5 лет, средний дебит скважины за первые 2 года эксплуатации, средний дебит скважины за первые 3 года эксплуатации, максимальный дебит скважины за последние 5 лет) на результаты обучения ННС.

Таблица 3 – Результаты корреляционного анализа

Входные параметры	Коэффициент корреляции с выходным параметром
gru	–0,03
pronic	–0,153
poris	–0,136
nasich	0,015
tolsh	0,02
plosh	0,028
bal1	0,01
bal2	0,007
podv1	0,017
podv2	0,017
kohvat	0,012
rnagn	0,013
debsos	0,19
obvod	–0,189

Анализ проведенных экспериментов позволил выявить ряд особенностей и закономерностей. Во-первых, при уменьшении порога корреляции с 0,15 до 0,05 и, соответственно, при увеличении числа входных параметров в модели точность классификации данных незначительно возрастает (с 90 до 90,53 % на обучающей выборке данных). Однако при этом существенно возрастает время построения модели (со 140 до 893 минут). Следовательно, не имеет смысла включать в модель большое число входных параметров. Для эффективной классификации данных достаточно использования четырех параметров.

Во-вторых, при изменении порога эффективности скважины относительно начального значения точность классификации незначительно снижается. Следовательно, экспертная оценка эффективности скважины на уровне 5 тонн в сутки является адекватной и согласуется с результатами экспериментов.

В-третьих, при подготовке данных к анализу лучшие результаты классификации достигаются при расчете среднего дебита нефти по скважине за первые 3 года эксплуатации. Соответственно, этот критерий наиболее адекватно (с точки зрения достигаемых результатов классификации) характеризует дебит нефти по скважине и, как следствие, оценку ее эффективности.

Таким образом, при формировании искомой базы знаний для подбора ГТМ (ввода скважин в эксплуатацию) на нефтяном месторождении реализованы следующие решения, отражающие оптимальный набор значений параметров:

- 1) число входных параметров: 4;
- 2) порог эффективности скважины: 5;
- 3) расчет эффективного дебита нефти (значения выходного параметра): средний дебит нефти по скважине за первые 3 года эксплуатации.

В результате анализа исходных данных и построения нейронечеткой модели сформирована база знаний для определения вариантов ГТМ (ввода скважин в эксплуатацию) на нефтяном месторождении. Рассмотрим фрагмент и характеристики сформированной базы знаний (табл. 4).

Таблица 4 – Фрагмент сформированной базы знаний

pronic	poris	debsos	obvod	result	CF
1	1	1	1	0	0,02
1	1	1	2	0	0,02
1	1	1	3	0	0,01
1	1	2	1	0	0,06
1	1	2	1	1	0,08
1	1	2	2	0	0,04
1	1	2	4	0	0,03
1	1	3	1	0	0,07
1	1	3	1	1	0,08
1	1	3	2	0	0,09
1	1	3	3	0	0,05
1	1	3	5	0	0,08

База знаний состоит из 448 нечетких правил. Каждое правило включает в себя следующие параметры:

- *pronic* – абсолютная проницаемость породы;
- *poris* – пористость породы;
- *debsos* – средний дебит соседних скважин;
- *obvod* – минимальная обводненность соседних скважин.

Значения входных параметров «1», «2», «3», «4», «5» означают, соответственно, «низкий», «ниже среднего», «средний», «выше среднего», «высокий».

Точность классификации данных при обучении ННС составила 97,54 %, точность классификации при тестировании составила 95,7 %, точность классификации при валидации модели – 95,68 %. Достигнутая точность классификации данных характеризует высокую аппроксимирующую способность сформированной базы знаний, а следовательно, ее адекватность и возможность практического использования.

Таким образом, полученные результаты точности классификации свидетельствуют о высокой обобщающей способности сформированной базы знаний.

Заключение. Описанные в данной работе нейронечеткая модель и программный комплекс прошли апробацию в управлении инвестиций и Центре обслуживания бизнеса ПАО «Татнефть», внедрены в бизнес-процесс формирования оптимальной программы геолого-технических мероприятий по всем объектам разработки в условиях ресурсных ограничений с учетом вариативности налоговых моделей. Использование результатов работы позволило:

- произвести выбор оптимальной стратегии проведения геолого-технических мероприятий по объектам разработки ПАО «Татнефть» в условиях ресурсных и финансовых ограничений;
- произвести ранжирование объектов разработки с целью рекомендации перехода на налоговую модель на дополнительный доход от добычи углеводородного сырья;
- снизить нагрузку на геолого-технологический персонал по рациональному выбору проектных скважин для бурения на нефтяных месторождениях.

Таким образом, описанный подход может быть использован при формировании баз знаний для оценки состояния объектов в различных предметных областях.

Библиографический список

1. Аждер, Т. Б. Системы поддержки принятия решений и информационные системы / Т. Б. Аждер, О. А. Гуреева // Уральский научный вестник. – 2019. – Т. 6, № 3. – С. 46–48.
2. Баринов, А. И. Использование модели нечетких нейронных сетей для формирования базы знаний по определению фишинговых сайтов / А. И. Баринов, Д. В. Катасёва, А. С. Катасёв // Вестник Технологического университета. – 2020. – Т. 23, № 10. – С. 64–67.
3. Башлыков, А. А. Роль человека-оператора как лица, принимающего решения, и элемента интеллектуальной системы управления сложными технологическими объектами / А. А. Башлыков // Автоматизация, телемеханизация и связь в нефтяной промышленности. – 2016. – № 12. – С. 10–18.
4. Глова, В. И. Преднастройка и оптимизация параметров нечеткой нейронной сети при формировании баз знаний экспертных систем / В. И. Глова, А. С. Катасёв, Г. С. Корнилов // Информационные технологии. – 2010. – № 5. – С. 15–19.
5. Гридина, Н. В. Построение гибридных нейронных сетей с использованием элементов нечеткой логики / Н. В. Гридина, И. А. Евдокимов, В. И. Солодовников // Искусственный интеллект и принятие решений. – 2019. – № 2. – С. 91–97.
6. Джуманов, О. И. Методы оптимизации идентификации нестационарных объектов на основе нейронечеткой сети с настройкой параметров вычислительных схем / О. И. Джуманов, С. М. Холмонов // Проблемы вычислительной и прикладной математики. – 2016. – № 3 (5). – С. 20–27.
7. Звонков, В. Б. Сравнительное исследование классических методов оптимизации и генетических алгоритмов / В. Б. Звонков, А. М. Попов // Вестник Сибирского государственного аэрокосмического университета им. академика М.Ф. Решетнева. – 2013. – № 4 (50). – С. 23–27.
8. Ключева, А. Р. Исследование и оценка состояния технических объектов на основе компьютерного моделирования / А. Р. Ключева // Инженерный вестник Дона. – 2018. – № 4 (51). – С. 120–132.
9. Комарцов, Л. Г. Исследование генетических алгоритмов для обучения многослойного персептрона / Л. Г. Комарцова, Д. С. Кадников // Нейрокомпьютеры: разработка, применение. – 2010. – № 12. – С. 12–19.
10. Коротеев, М. В. Формы функции принадлежности лингвистических переменных экономических показателей / М. В. Коротеев // Аудит и финансовый анализ. – 2012. – № 2. – С. 239–244.
11. Кравченко, Т. К. Создание систем поддержки принятия решений: интеграция преимуществ отдельных подходов / Т. К. Кравченко, Н. Н. Середенко // Искусственный интеллект и принятие решений. – 2012. – № 1. – С. 39–47.
12. Кузовлев, В. И. Повышение качества данных с использованием методики поиска аномалий на примере портала открытых данных правительства Москвы / В. И. Кузовлев, А. О. Орлов // Инженерный вестник. – 2014. – № 8. – С. 7–14.

13. Назаров, А. О. Выбор вида функции принадлежности в нечеткой модификации алгоритма COBWEB для задачи формирования пользовательских ролей / А. О. Назаров, И. В. Аникин // Вестник Казанского государственного технического университета им. А.Н. Туполева. – 2014. – № 2. – С. 214–219.

14. Пантелеев, А. В. Градиентные методы оптимизации в машинном обучении идентификации параметров динамических систем / А. В. Пантелеев, А. В. Лобанов // Моделирование и анализ данных. – 2019. – № 4. – С. 88–99.

15. Сафронов, В. В. Методы построения функций принадлежности / В. В. Сафронов, Ю. В. Ведерников // Информационные технологии. – 2007. – № 11. – С. 8–11.

References

1. Azhder, T. B., Gureeva, O. A. Sistemy podderzhki prinyatiya resheniy i informatsionnye sistemy [Decision support systems and information systems]. *Uralskiy nauchnyy vestnik* [Ural Scientific Bulletin], 2019, vol. 6, no. 3, pp. 46–48.

2. Barinov, A. I., Katasyova, D. V., Katasyov, A. S. Ispolzovanie modeli nechetkikh neyronnykh setey dlya formirovaniya bazy znaniy po opredeleniyu fishingovykh saytov [Using a fuzzy neural network model to form a knowledge base for identifying fishing sites]. *Vestnik Tekhnologicheskogo universiteta* [Bulletin of Technological University], 2020, vol. 23, no. 10, pp. 64–67.

3. Bashlykov, A. A. Rol cheloveka-operatora kak litsa, prinimayushchego resheniya, i elementa intellektualnoy sistemy upravleniya slozhnyimi tekhnologicheskimi obektami [The role of a human operator as a decision maker and an element of an intelligent control system for complex technological objects]. *Avtomatizatsiya, telemekhanizatsiya i svyaz v neftyanoy promyshlennosti* [Automation, Telemechanization and Communication in the Oil Industry], 2016, vol. 12, pp. 10–18.

4. Glova, V. I., Katasyov, A. S., Kornilov, G. S. Prednastroyka i optimizatsiya parametrov nechetkoy neyronnoy seti pri formirovaniy baz znaniy ekspertnykh sistem [Presetting and optimization of fuzzy neural network parameters when forming knowledge bases of expert systems]. *Informatsionnye tekhnologii* [Information Technologies], 2010, no. 5, pp. 15–19.

5. Gridina, N. V., Evdokimov, I. A., Solodovnikov, V. I. Postroenie gibridnykh neyronnykh setey s ispolzovaniem elementov nechetkoy logiki [Constructing hybrid neural networks using fuzzy logic elements]. *Iskusstvennyy intellekt i prinyatie resheniy* [Artificial intelligence and decision making], 2019, no. 2, pp. 91–97.

6. Dzhumanov, O. I., Kholmonov, S. M. Metody optimizatsii identifikatsii nestatsionarnykh obektov na osnove neyro-nechetkoy seti s nastroykoy parametrov vychislitelnykh skhem [Methods for optimizing the identification of non-stationary objects based on a neuro-fuzzy network with setting the parameters of computational circuits]. *Problemy vychislitelnoy i prikladnoy matematiki* [Problems of Computational and Applied Mathematics], 2016, no. 3 (5), pp. 20–27.

7. Zvonkov, V. B., Popov, A. M. Sravnitelnoe issledovanie klassicheskikh metodov optimizatsii i geneticheskikh algoritmov [Comparative study of classical optimization methods and genetic algorithms]. *Vestnik Sibirskogo gosudarstvennogo aerokosmicheskogo universiteta imeni akademika M.F. Reshetneva* [Bulletin of the Siberian State Aerospace University named after Academician M.F. Reshetnev], 2013, no. 4 (50), pp. 23–27.

8. Klyueva, A. R. Issledovanie i otsenka sostoyaniya tekhnicheskikh obektov na osnove kompyuternogo modelirovaniya [Technical objects state research and assessment based on computer simulation]. *Inzhenernyy vestnik Dona* [Engineering Bulletin of the Don], 2018, no. 4 (51), pp. 120–132.

9. Komartsov, L. G., Kadnikov, D. S. Issledovanie geneticheskikh algoritmov dlya obucheniya mnogosloynogo perseptrona [Study of genetic algorithms for training multilayer perceptron]. *Neyrokomp'yutery: razrabotka, primeneniye* [Neurocomputers: Development, Application], 2010, no. 12, pp. 12–19.

10. Koroteev, M. V. Formy funktsii prinaldezhnosti lingvisticheskikh peremennykh ekonomicheskikh pokazateley [Forms of the membership function of linguistic variables of economic indicators]. *Audit i finansovyy analiz* [Audit and Financial Analysis], 2012, no. 2, pp. 239–244.

11. Kravchenko, T. K., Seredenko, N. N. Sozdanie sistem podderzhki prinyatiya resheniy: integratsiya preimushchestv ot delnykh podkhodov [Constructing decision support systems: integrating the benefits of separate approaches]. *Iskusstvennyy intellekt i prinyatie resheniy* [Artificial intelligence and decision making], 2012, no. 1, pp. 39–47.

12. Kuzovlev, V. I., Orlov, A. O. Povyshenie kachestva dannykh s ispolzovaniem metodiki poiska anomalii na primere portala otkrytykh dannykh pravitelstva Moskvy [Improving the quality of data using the anomaly search technique on the example of the open data portal of the Moscow government]. *Inzhenernyy vestnik* [Engineering Bulletin], 2014, no. 8, pp. 7–14.

13. Nazarov, A. O., Anikin, I. V. Vybor vida funktsii prinaldezhnosti v nechetkoy modifikatsii algoritma COBWEB dlya zadachi formirovaniya polzovatelskikh roley [Choosing the type of membership function in the fuzzy modification of the COBWEB algorithm for user roles generating problem]. *Vestnik Kazanskogo gosudarstvennogo tekhnicheskogo universiteta imeni A.N. Tupoleva* [Bulletin of the Kazan State Technical University named after A.N. Tupolev], 2014, no. 2, pp. 214–219.

14. Panteleev, A. V., Lobanov, A. V. Gradientnye metody optimizatsii v mashinnom obuchenii identifikatsii parametrov dinamicheskikh sistem [Gradient optimization methods in machine learning for identification of dynamic system parameter]. *Modelirovanie i analiz dannykh* [Modeling and Data Analysis], 2019, no. 4, pp. 88–99.

15. Safronov, V. V., Vedernikov, Yu. V. Metody postroeniya funktsiy prinaldezhnosti [Methods of membership functions constructing]. *Informatsionnye tekhnologii* [Information Technologies], 2007, no. 11, pp. 8–11.