

УДК 004.89

**АНАЛИЗ АЛГОРИТМИЧЕСКИХ МЕТОДОВ ФОРМИРОВАНИЯ ТЕРМОВ ЛИНГВИСТИЧЕСКИХ ПЕРЕМЕННЫХ***Е. Н. Чуйкова., И. Р. Кикоть*

Донской государственной технической университет, Ростов-на-Дону, Российская Федерация  
[elenchus@mail.ru](mailto:elenchus@mail.ru)

Проведен анализ существующих алгоритмических методов формирования функций принадлежности термов лингвистических переменных, описаны их основные свойства, преимущества и недостатки. На основе анализа сформулированы выводы о возможности применения рассмотренных методов; определены направления дальнейших исследований в области создания алгоритмов формирования функций принадлежности для организации нечеткого поиска в реляционной базе данных; выявлены характеристики, которыми должны обладать разрабатываемые алгоритмы; выбраны методы решения поставленных задач.

**Ключевые слова:** нечеткая система, лингвистическая переменная, нечеткое множество, функция принадлежности, кластеризация, генетический алгоритм, нейронная сеть.

**Введение.** Важнейшей задачей, от решения которой зависит качество функционирования нечетких систем, является формирование термов лингвистических переменных (ЛП), используемых в данной предметной области. В данном случае предполагается определение числа термов (нечетких множеств), их размещения на базовой шкале, формы функций принадлежности (треугольная, трапециевидная, S-функция и т. д.). Функции принадлежности традиционно строят на основе экспертной информации, что затрудняет автоматизацию процесса их формирования, поэтому усилия многих исследователей направлены на создание алгоритмических методов их построения. В настоящее время предложено несколько алгоритмических методов определения функций принадлежности. В данной работе проводится их анализ.

**Постановка задачи.** Существуют приложения нечеткой логики (например, организация нечеткого поиска в реляционной базе данных), требующие простых алгоритмов формирования

UDC 004.89

**ANALYSIS OF ALGORITHMIC METHODS OF LINGUISTIC VARIABLE TERMS FORMATION***E. N. Chuykova, I.R. Kikot*

Don State Technical University, Rostov-on-Don, Russian Federation  
[elenchus@mail.ru](mailto:elenchus@mail.ru)

The article presents the analysis of existing algorithmic construction methods of membership functions of linguistic variables terms, their main properties, advantages and disadvantages. Based on the analysis the conclusions on the considered methods application are made; areas for further research in the field of creation of algorithms of membership function formation for fuzzy search organization in a relational database are identified; the characteristics that developed algorithms should possess are identified; problem solution methods are selected.

**Keywords:** fuzzy system, linguistic variable, fuzzy set, the membership function, clustering, genetic algorithm, neural network.

функций принадлежности. Такие алгоритмы должны обеспечивать оперативный расчет значений функций принадлежности средствами языков данных (например, *SQL*), не обладающих большими вычислительными возможностями, присущими универсальным языкам программирования [1]. Расчет должен осуществляться в процессе обработки запроса пользователя к базе данных и не требовать больших затрат времени, чтобы не принуждать пользователя к длительному ожиданию ответа системы. Указанные алгоритмы должны определять термы лингвистических переменных по выборке экспериментальных данных, полученных, например, из таблиц базы данных. Следует проанализировать существующие методы автоматического построения функций принадлежности, выявить степень их применимости в решении задачи организации нечеткого поиска в реляционной базе данных и выработать рекомендации по использованию полученных результатов анализа.

**Методы, основанные на алгоритмах кластеризации.** В [2, 3] описаны методы, использующие нечеткий алгоритм *c*-средних для определения функций принадлежности. Метод нечеткой кластеризации *c*-средних позволяет разбить имеющееся множество точек (объектов) на заданное число нечетких множеств (кластеров), в которых объекты более схожи между собой, чем с объектами из других кластеров. Особенностью метода является использование нечеткой матрицы принадлежности  $U$  с элементами  $u_{ij}$ , определяющими принадлежность  $i$ -го элемента исходного множества объектов —  $j$ -му кластеру. Кластеры описываются своими центрами  $c_j$  — точками того же пространства, которому принадлежит исходное множество точек. Нечеткие методы кластеризации позволяют одному и тому же объекту принадлежать одновременно нескольким (или даже всем) кластерам, но с различной степенью. В ходе нечеткой кластеризации решается задача минимизации целевой функции:

$$E = \sum \sum u_{ij}^m \cdot \|x_j - c_j\|^2 \text{ при ограничениях } \sum_j u_{ij} = 1, i = 1..p,$$

где  $m$  — экспоненциальный вес, определяющий нечеткость кластеров;  $p$  — число кластеров.

Функции принадлежности нечетких кластеров соответствуют искомым функциям принадлежности нечетких множеств.

Для определения оптимальной формы функции принадлежности, ассоциируемой с каждым термом, используется критерий среднеквадратичной ошибки.

Метод чувствителен к выбросам и предполагает, что число нечетких множеств заранее определено.

**Методы, основанные на генетических алгоритмах.** Генетические алгоритмы являются процедурами оптимизации некоторой целевой функции, основанными на принципах естественной эволюции. В них сгенерированная случайным образом начальная популяция решений подвергается воздействию генетических операций, таких как селекция, скрещивание и мутация. В результате

получается новая популяция решений с улучшенными основными свойствами (значениями целевой функции). Данный процесс продолжается итеративно до достижения оптимума.

В [4] предложен генетический алгоритм для определения числа термов ЛП на основе двух критериев — мощности нечетких множеств и требуемых затрат времени для их определения. Генетический алгоритм реализует многокритериальную оптимизацию на основе концепции Парето. Рассматриваются нечеткие множества треугольной формы.

В [5] описан генетический алгоритм, используемый для определения типа и мощности терм-множеств лингвистических переменных. В структуре хромосомы присутствуют гены, кодирующие:

- мощности терм-множеств лингвистической переменной;
- коэффициенты, характеризующие пересечения носителей функций принадлежности соответствующей лингвистической переменной;
- тип функций принадлежности нечетких множеств.

В качестве критерия оптимальности используется функция максимума вероятности вхождения объектов в различные терм-множества лингвистической переменной при минимальной мощности терм-множества.

Подобно алгоритмам кластеризации недостатком генетических алгоритмов является необходимость предварительного задания числа нечетких множеств.

**Методы, основанные на нейронных сетях.** Для определения функций принадлежности используется аппарат нейронных сетей. В [6] описана процедура извлечения знаний для определения структуры и параметров базы нечетких правил посредством двухфазного обучения нейронной сети. Процедура также включает предварительный выбор свойств нечетких множеств для получения более простых и надежных нечетких правил и последующую обработку для формирования выходов полученных нечетких правил. Предложенная структура представляет собой объединение процедуры извлечения знаний и двух дополнительных фаз обработки. Процедура извлечения знаний, ядро структуры, реализуется посредством обучения нечеткой нейросети, в структуре которой извлекаемые знания кодируются в виде нечетких правил. Этапы выбора переменных и обработки выходов предназначены для повышения точности и надежности формируемой базы нечетких правил, хотя эти этапы не обязательны и в некоторых случаях могут быть опущены.

**Методы, основанные на использовании специальных мер.** В [7] для определения функций принадлежности используется понятие энтропии. Согласно теории информации энтропия измеряет степень неопределенности информационной системы. Большая энтропия информационной системы определяет больший объем информации, содержащейся в системе. Аналогично, большая энтропия в появлении нечеткого события определяет больший объем информации, связанной с

нечетким событием. Таким образом, метод отыскивает функцию принадлежности, которая максимизирует энтропию нечеткого события. Используется функция принадлежности типа  $S$ -функции.

В [8] предложен метод определения нечетких множеств на основе двух параметров — нечеткой энтропии и индекса нечеткости. Понятие нечеткой энтропии определяется как количество нечеткой информации, содержащейся в нечетком множестве или нечеткой системе. Математическое определение индекса нечеткости приведено в [8]. Метод направлен на поиск функции, максимизирующей количество информации об объекте. Так как энтропия — мера информации, то функция, максимизирующая энтропию, наиболее информативна. Для поиска функции, наилучшим образом описывающей нечеткое множество, авторы используют индекс нечеткости. Результатом является функция, имеющая наибольший индекс нечеткости. Таким образом, исходная задача заменяется задачей поиска функции, оптимальной в отношении общей энтропии и индекса нечеткости. Используется  $S$ -образная функция принадлежности. Подход применен в области распознавания образов.

**Другие подходы.** Предложено несколько других методов для автоматического определения нечетких множеств. Эвристические методы, использующие predetermined формы нечетких множеств, дают хорошие результаты при решении задач классификации в области машинного зрения [9].

В [10] для определения функций принадлежности нечетких множеств используются гистограммы.

Во многих алгоритмах число нечетких множеств определяется эмпирически в диапазоне 2–7 и выполняется равномерное распределение нечетких множеств на базовой шкале.

**Выводы.** Существует несколько алгоритмических методов, допускающих автоматическое построение функций принадлежности нечетких множеств, использующих различные подходы, но нет универсальных правил, рекомендаций и даже единого мнения в вопросе выбора лучшего решения, применимого в любой области. Выбор метода построения функции принадлежности зависит от области применения и личных предпочтений исследователя. В результате при решении конкретной задачи приходится тестировать множество методов, чтобы найти наиболее подходящий в данной ситуации.

Функции принадлежности часто задаются эмпирически. Трапецевидная и треугольная формы получили наиболее широкое распространение, поскольку дают хорошие результаты для многих областей применения. Большинство авторов предлагают равномерное размещение нечетких множеств в области определения.

Большинство предлагаемых методов характеризуются высокой вычислительной сложностью.

Недостатком генетических алгоритмов также является большое время сходимости, зависимость от способа кодирования хромосом.

Нейронные сети требуют настройки параметров, формирования обучающей выборки и нечеткой модели «входы — выход». Функция принадлежности чувствительна к обучающей выборке и структуре нечеткой модели, т. е. к базе знаний.

Использование индексов дает хорошие результаты, но соответствующие методы не позволяют определить оптимальное число нечетких множеств.

Функции принадлежности, формируемые с помощью рассмотренных методов, не всегда удовлетворяют предъявляемым к ним требованиям [11] и часто нуждаются в дополнительной аппроксимации.

Во многих случаях наилучшим решением остается эмпирическое определение функций принадлежности.

Для реализации некоторых методов используется специализированное программное обеспечение.

Таким образом, для решения задач организации нечеткого поиска в реляционной базе данных целесообразно разработать простые алгоритмы, отличающиеся невысокой вычислительной сложностью и не требующие больших затрат времени на решение задачи. Алгоритмы должны выполнять автоматическое построение функций принадлежности на основе экспериментальных данных с использованием статистических методов и эвристик, позволяющих избежать высокой сложности существующих методов, зависящих главным образом не от области применения, а от структуры набора данных, допускающих трансляцию шагов алгоритма в стандартные конструкции языка *SQL*.

#### Библиографический список

1. Чуйкова, Е. Н. Реализация нечеткого выбора оборудования в системе проектирования информационной сети / Е. Н. Чуйкова // Вестник Дон. гос. техн. ун-та. — 2014. — Т. 14, № 3 (78). — С. 164–171.
2. Chen, M.-S. Fuzzy clustering analysis for optimizing membership functions / M.-S. Chen, S.-W. Wang / Fuzzy Sets and Systems. — 1999. — Vol. 103. — P. 239–254.
3. Liao, T.-W. A fuzzy c-means variant for the generation of fuzzy term sets / T.-W. Liao, A.-K. Celmins, R.-J. Hammell I. Fuzzy Sets and Systems. — 2001. — Vol. 135 (2). — P. 241–257.
4. Kaya, M. Integrating multi-objective genetic algorithms into clustering for fuzzy association rules mining / M. Kaya, R. Alhajj // Proceedings of the Fourth IEEE International Conference on Data Mining / IEEE Computer Society. — Washington, 2004. — P. 431–434.

5. Панфилов, С. А. Генетический алгоритм оптимизации структуры лингвистических переменных при построении баз знаний нечетких систем [Электронный ресурс] / С. А. Панфилов, А. В. Язенин // Программные продукты и системы. — 2004. — № 1. — Режим доступа: <http://www.swsys.ru/index.php?page=article&id=599/> (дата обращения: 11.05.16).
6. Knowledge discovery by a neuro-fuzzy modeling framework / G. Castellano [et al.] // Fuzzy Sets and Systems. — 2005. — Vol. 149. — P. 187–207.
7. Cheng, H.-D. Automatically determine the membership function based on the maximum entropy principle / H.-D. Cheng, J.-R. Chen // Information Sciences. — 1997. — Vol. 96 (3/4). — P. 163–182.
8. Nieradka, G. A method for automatic membership function estimation based on fuzzy measures / G. Nieradka, B. Butkiewicz // International Fuzzy Systems Association World Congress — IFSA. — 2007. — P. 451–460.
9. Ishibuchi, H. Efficient fuzzy partition of pattern space for classification problems / H. Ishibuchi, K. Nozaki, H. Tanaka // Fuzzy Sets and Systems. — 1993. — Vol. 59 (3). — P. 295–304.
10. Medasani, S. An overview of membership function generation techniques for pattern recognition / S. Medasani, J. Kim, R. Krishnapuram // International Journal of Approximate Reasoning. — 1998. — Vol. 19. — P. 391–417.
11. Базы данных. Интеллектуальная обработка информации / В. В. Корнеев [и др.]. — Москва : Издатель Молгачева С. В. ; Нолидж, 2001. — 496 с.