

УДК 519.87

Doi: 10.31772/2712-8970-2022-23-1-33-42

Для цитирования: Плешкова Т. С., Становов В. В. Сравнение методов кластеризации данных для автоматического определения грануляции в генетической нечеткой системе // Сибирский аэрокосмический журнал. 2022. Т. 23, № 1. С. 33–42. Doi: 10.31772/2712-8970-2022-23-1-33-42.

For citation: Pleshkova T. S., Stanovov V. V. Comparison of data clustering methods for automatic determination of granulation in a genetic fuzzy system. *Siberian Aerospace Journal*. 2022, Vol. 23, No. 1, P. 33–42. Doi: 10.31772/2712-8970-2022-23-1-33-42.

Сравнение методов кластеризации данных для автоматического определения грануляции в генетической нечеткой системе*

Т. С. Плешкова¹, В. В. Становов

Сибирский федеральный университет
Российская Федерация, 660041, г. Красноярск, просп. Свободный, 79

¹E-mail: tatyana.pleshkova2310@gmail.com

В статье предлагается применение методов кластеризации для определения наиболее подходящего количества нечетких термов при построении генетической нечеткой системы. При этом система на нечеткой логике применяется для решения задач классификации данных и автоматически формируется генетическим алгоритмом. В работе использовался генетический алгоритм с кодировкой термов и классов в бинарную строку, при этом каждый индивид кодировал базу правил. Для построения базы правил необходимо задавать такой параметр, как количество нечетких термов, так как он существенно влияет на качество сформированных классификаторов. Для выявления лучшего метода кластеризации данных было проведено сравнение наиболее известных алгоритмов: DBSCAN, k -средних и алгоритм среднего сдвига. Для оценки эффективности выбранного количества нечетких термов были проведены вычислительные эксперименты на нескольких наборах данных. По результатам было определено, что алгоритм среднего сдвига подбирает такое количество термов, которое позволяет строить более точные классификаторы в сравнении с двумя другими методами, участвовавшими в тестировании. Также было проведено сравнение с альтернативными методами классификации, такими как k ближайших соседей, метод опорных векторов и нейронные сети, в результате которого предложенный метод показал сравнимое качество классификации. Разработанный подход к автоматизации определения количества термов позволяет исключить ручной подбор грануляции для различных данных, снижая затраты на создание эффективной нечеткой системы для задачи классификации.

Ключевые слова: классификация, нечеткая логика, генетический алгоритм, DBSCAN, k -средних, алгоритм среднего сдвига.

Comparison of data clustering methods for automatic determination of granulation in a genetic fuzzy system

T. S. Pleshkova¹, V. V. Stanovov

Siberian Federal University
79, Svobodny Av., Krasnoyarsk, 660041, Russian Federation

¹E-mail: tatyana.pleshkova2310@gmail.com

* Работа выполнена в рамках программы государственной поддержки ведущих научных школ (грант Президента РФ НШ-421.2022.4).

The work was carried out within the framework of the state support program for leading scientific schools (grant of the President of the Russian Federation NSH-421.2022.4).

The paper proposes applying clustering methods to determine the most appropriate number of fuzzy terms to design a genetic fuzzy system. The fuzzy logic system in this study is used to solve data classification problems and is automatically generated by a genetic algorithm. In this study we used a genetic algorithm with encoding of terms and classes into a binary string, while each individual encoded a rule base. It is necessary to set the number of fuzzy terms parameter to build a rule base, since it significantly affects the quality of the generated classifiers. To determine the best method of data clustering, the most well-known algorithms were compared: DBSCAN, k-means and the mean shift algorithm. To evaluate the efficiency of the selected number of fuzzy terms, computational experiments were performed on several data sets. Based on the results, it was determined that the mean shift algorithm selects such a number of terms that allows building more accurate classifiers in comparison to the other two methods involved in testing. A comparison was also made with alternative classification methods such as k nearest neighbors, support vector machine and neural networks, as a result of which the proposed method showed comparable classification quality. The developed approach to automating the determination of the number of terms makes it possible to exclude manual selection of granulation for various data, reducing the cost of creating an effective fuzzy system for solving classification problems.

Keywords: classification, fuzzy logic, genetic algorithm, DBSCAN, k-means, Mean Shift.

Введение

Одним из важных инструментов для анализа данных является классификация, суть которой состоит в группировке объектов по классам в соответствии с их общими признаками. Известно, что классификация – это необходимый шаг в понимании любой области исследования [1]. На данный момент существует множество методов, позволяющих решать задачи классификации. Но многие современные и популярные методы являются своего рода черным ящиком [2], который получает на вход данные и выводит ответ, но проводит классификацию без предоставления информации о происходящих внутри процессах.

Не все задачи допускают использование черных ящиков, так как часто, в зависимости от области исследования, необходимо понимать, какие признаки повлияли на результат классификации. Интерпретируемая классификация применяется для многих задач, в которых важно обеспечить возможность обосновывать решения и объяснять результат, в особенности для таких задач, в которых цена ошибки слишком велика. К таким задачам можно отнести классификацию различной аппаратуры космического применения. Поэтому при изучении различных методов, используемых для классификации данных, было принято решение использовать эволюционные алгоритмы для построения систем на нечеткой логике (НЛ) [3].

Нечеткая логика была предложена американским ученым математиком Лотфи Заде с целью описания расплывчатых определений естественного языка [4]. Подход НЛ имитирует способ принятия решений человеком и вместо четких понятий истина (1) и ложь (0) включает все промежуточные значения между 0 и 1. В нечеткой классификации рассматривается отнесение каждого объекта ко всем классам с определенными степенями принадлежности, которые в свою очередь указывают силу ассоциации между объектом и соответствующим классом [5]. Поэтому одним из важных этапов является подбор нечеткой базы правил, представляющей собой набор независимых правил, каждое из которых выражает причинно-следственную связь между входными переменными и соответствующим классом с помощью нечеткого множества или термов [6].

При этом существует задача подбора нечеткой грануляции (термов), так как количество термов зависит от данных. Решение данной задачи является многоуровневым процессом, поэтому в статье речь пойдет о создании базы правил с использованием генетического алгоритма, подборе нечеткой грануляции с помощью кластеризации данных, описании экспериментов, результатах и направлении будущих исследований.

Построение базы правил генетическим алгоритмом

Как уже было сказано ранее, качество подобранной базы правил (БП) влияет на точность классификации (процент верно классифицированных объектов с помощью подобранной БП) и

интерпретируемость, поэтому в разработанном алгоритме использован поиск БП с помощью генетического алгоритма (ГА) [7]. Стоит заметить, что гибридизация ГА и НЛ уже рассматривалась ранее, например, в работе [6] предложен самонастраивающийся эволюционный алгоритм формирования нечетких систем для решения задач классификации с представлением баз правил в форме матриц переменной размерности. В работе [8] разработан алгоритм, который представляет собой гибридизацию Питтсбургского и Мичиганского подходов, которая реализована в рамках эволюционной многокритериальной оптимизации.

Генетические алгоритмы входят в состав эволюционных алгоритмов поиска, которые объединяют различные варианты использования эволюционных принципов для достижения поставленной цели. ГА позволяют осуществлять поиск оптимального решения для широкого списка прикладных решений. ГА делится на несколько этапов: инициализация популяции, селекция, скрещивание, мутация и формирование нового поколения [9]. Процесс поиска начинается с набора индивидов или же с популяции. Индивид характеризуется набором параметров (переменных), известных как гены. Гены объединяются в цепочку, образуя хромосому. Размер хромосомы может меняться в зависимости, как например, в данной работе, от количества термов или классов. Определить количество битов на хромосому можно по формуле (1)

$$len = (var \cdot bt + bc) \cdot m, \tag{1}$$

где var – количество переменных в выборке данных; bt – количество битов, необходимое на кодирование одного терма; bc – количество битов, необходимое на кодирование одного класса; m – количество правил.

Графическое представление кодирования правила в хромосому с тремя термами представлено на рис. 1.

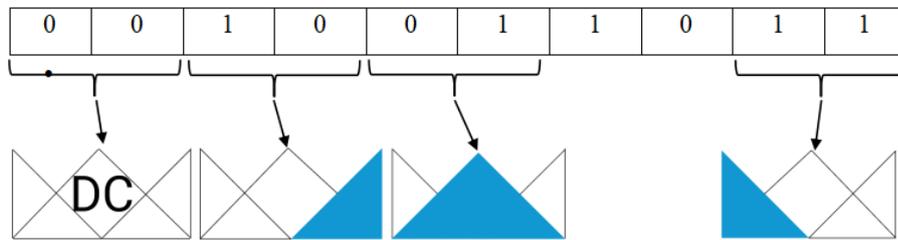


Рис. 1. Графическое представление кодирования правила в хромосому

Fig. 1. Graphical representation of encoding a rule into a chromosome

Стоит заметить, что в работе рассматривалось различное количество термов, а именно от 2 до 7 (рис. 2).

Изначально проводится создание начальной популяции случайным образом, при этом заполняется массив, имеющий размерность количество индивидов N , умноженное на len . После формирования и заполнения необходимо произвести перевод бинарного массива в фенотип, т. е. правило, представляющее собой последовательность целых чисел, кодирующих номера термов. В дальнейшем эти правила используются для оценки точности БП.

Перевод осуществляется сначала для битов, выделенных для переменных, т. е. рассматриваются каждые bt позиций в строке длиной len . При переводе стоит учитывать дополнительный терм – игнорирование, который позволяет упростить правила и саму БП в целом. А затем перевод осуществляется для последних bc битов, выделенных для класса.

После перевода с помощью функции пригодности определяется качество индивида, т. е. точность БП. Определение класса происходит с помощью формулы (2) (определение правила-победителя). Чтобы определить правило-победитель, необходимо для каждого из правил рас-

считать минимум по степеням принадлежности по всем переменным [6]. Правило-победитель будет определяться с максимальной степенью принадлежности

$$N_{Rule} = \arg \max_m (\min_f (\mu_{j_m, f}(x_f))), \quad (2)$$

где N_{Rule} – номер правила-победителя; m – число правил; $j_{m,f}$ – элемент правила; μ – степень принадлежности; x – переменная с индексом f [6].

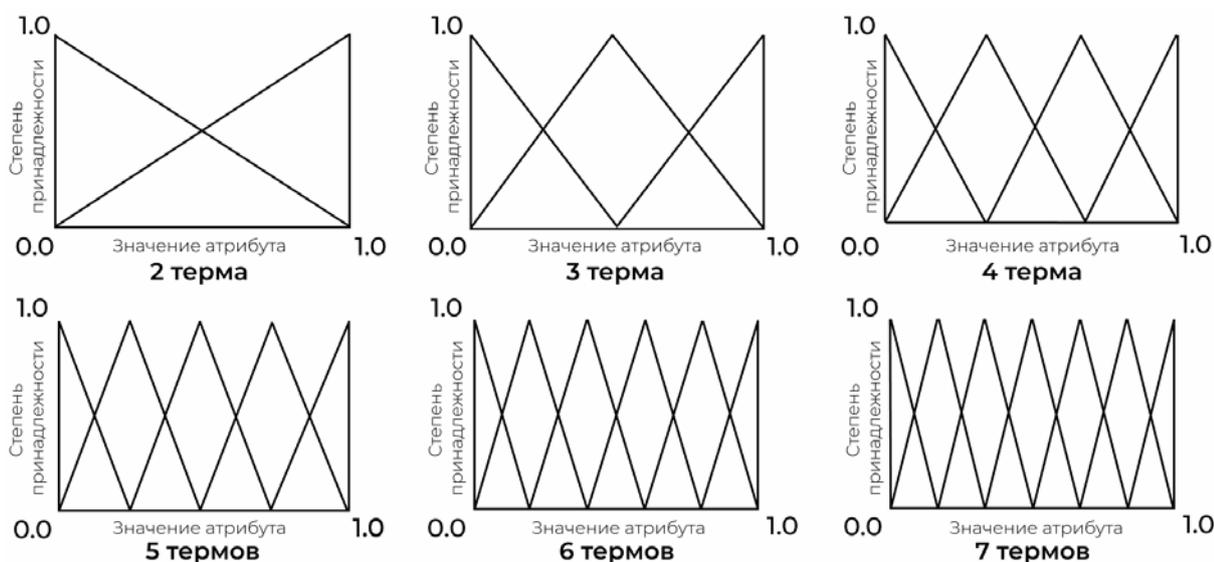
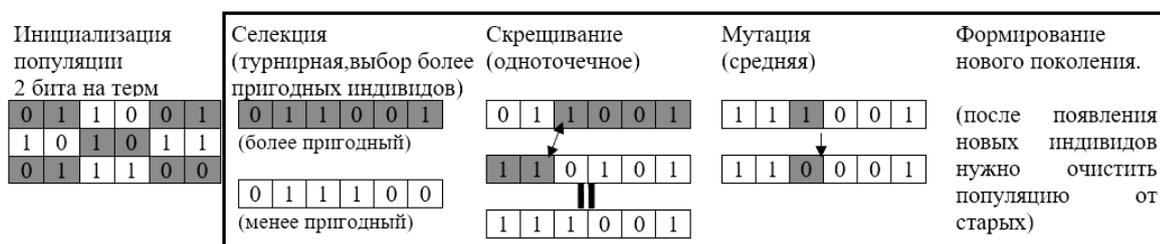


Рис. 2. Визуальное отображение различного количества термов

Fig. 2. Visual display of different number of term

После получения оценки пригодности каждого индивида можем определить вероятность выбора какого-то конкретного индивида для воспроизводства. На основании собственных вычислений и анализа работ [6; 10; 11] в качестве параметров ГА были выбраны турнирная селекция с $T=2$, одноточечное скрещивание и средняя мутация. Структурная схема ГА представлена на рис. 3.



Этапы повторяются до тех пор, пока результат не будет удовлетворительным или не будет выполнено условие, что пройдено максимальное количество поколений.

Рис. 3. Структурная схема генетического алгоритма

Fig. 3. Block diagram of the genetic algorithm

Стоит отметить, что на эффективность работы ГА сильно влияет конфигурация его параметров, выбор которых зависит от решаемой задачи, поэтому определение нужного набора является дополнительным исследованием.

Тестирование алгоритма

Тестирование производилось на нормированных данных с репозитория UCI Machine Learning [12]. Описание характеристик выборок представлено в табл. 1.

Таблица 1

Характеристики тестовых наборов данных

Выборка	Число измерений	Число переменных	Число классов
Australian credit	690	14	2
Iris	150	4	3
Seeds	210	7	3
Glass	214	9	6
Ionosphere	351	34	2

В процессе тестирования алгоритма подбирались следующие параметры: количество правил, количество термов. Подбор осуществлялся перебором, был реализован цикл по термам, в котором был цикл по правилам. Алгоритм запускался при следующих параметрах: количество индивидов – 100, процент разбиения выборки – 80/20, количество поколений – 800, разбиение было случайным, не стратифицированным. После каждых ста поколений замерялась точность, т. е. процент верно классифицированных объектов на тестовой выборке, которая не участвовала в процессе обучения. Точность замеряется для анализа качества процесса работы алгоритмы. По 10 запускам подсчитывалось среднее значение. С результатами классификации на выборке Ionosphere можно ознакомиться в табл. 2, 3 для двух и трех термов соответственно.

Таблица 2

Результаты работы алгоритма классификации данных на выборке «Ionosphere», 2 термина

Кол-во правил \ Кол-во поколений	3	4	5	6	7	8	9	10	11	12
300	0,765	0,787	0,762	0,772	0,768	0,746	0,763	0,737	0,786	0,772
400	0,765	0,787	0,763	0,776	0,765	0,746	0,763	0,735	0,785	0,773
500	0,768	0,790	0,763	0,776	0,765	0,744	0,761	0,735	0,786	0,772
600	0,769	0,790	0,763	0,776	0,765	0,744	0,759	0,735	0,787	0,772
700	0,772	0,792	0,766	0,775	0,779	0,744	0,759	0,735	0,787	0,770
800	0,772	0,792	0,766	0,777	0,779	0,744	0,759	0,735	0,790	0,770

Таблица 3

Результаты работы алгоритма классификации данных на выборке «Ionosphere», 3 термина

Кол-во правил \ Кол-во поколений	3	4	5	6	7	8	9	10	11	12
300	0,548	0,546	0,580	0,573	0,566	0,559	0,572	0,600	0,576	0,577
400	0,548	0,546	0,580	0,575	0,565	0,562	0,570	0,600	0,575	0,579
500	0,549	0,546	0,580	0,575	0,565	0,561	0,570	0,600	0,576	0,579
600	0,549	0,546	0,585	0,583	0,565	0,559	0,573	0,600	0,576	0,579
700	0,549	0,548	0,585	0,587	0,569	0,559	0,577	0,600	0,576	0,579
800	0,551	0,546	0,585	0,590	0,573	0,568	0,577	0,600	0,577	0,579

Анализируя табл. 2 и 3, можно сделать вывод, что наибольшая точность классификации достигается при разных параметрах, т. е. существует зависимость точности классификации от количества термов. Для решения проблемы определения наилучшего количества термов могут быть использованы различные алгоритмы, но ниже в статье проведено сравнение методов кластеризации данных для автоматического определения количества термов.

Автоматическое определение нечеткой грануляции

Определение нечеткой грануляции можно решить различными способами. В данной статье данная задача решена с помощью кластеризации данных, поэтому проведено сравнение следующих методов:

1. Метод DBSCAN (Density-based spatial clustering of applications with noise), вариация данного алгоритма была рассмотрена в [13]. Параметр „число соседей“ для различных наборов данных подсчитывался, как половина от числа объектов класса, содержащего наименьшее количество объектов. Параметр „радиус ε -окрестности“ генерировался по нормальному распределению со средним значением равным среднему евклидову расстоянию по одной переменной, а стандартное отклонение равным стандартному отклонению расстояний по этой же переменной.

2. Метод k-средних с использованием в качестве метрики оценки (критерием качества оценки кластеризации) метода локтя [14]. Метод локтя основывается на нахождении искажения, т. е. суммы квадрата расстояния (SSE) между точками данных и центроидами назначенных им кластеров [15]. Для определения наилучшего количества термов в алгоритме сделан перебор чисел от 2 до 7 и оценка того, насколько хорошо различные k кластеры подходят для выборки.

3. Mean Shift [16] – непараметрический алгоритм, используемый в обучении без учителя. Преимущество данного алгоритма в том, что не нужно заранее указывать количество кластеров, так как алгоритм среднего сдвига будет определять число сразу по данным.

Результаты экспериментов представлены на рис. 4 и табл. 4, где жирным шрифтом выделены лучшая точность и количество термов для различных данных.

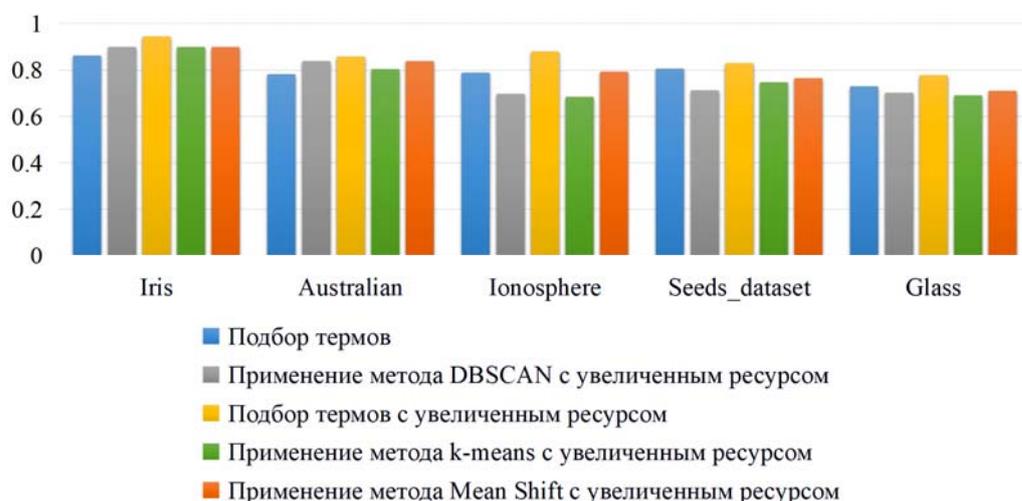


Рис. 4. Визуализация результатов сравнения методов кластеризации

Fig. 4. Visualization of results of comparison of clustering methods

Изначально подбор термов осуществлялся на 100 индивидах. При тестировании количество индивидов было увеличено до 300, что дало более высокие результаты. На рис. 4 данное улучшение обозначено как подбор термов с увеличенным ресурсом.

По результатам, представленным в табл. 4 и на рис. 4, можно сделать вывод, что метод Mean Shift точнее предсказывает количество термов для различного набора данных. Методы DBSCAN и k-средних уступают на некоторых выборках.

С результатами лучших конфигураций работы алгоритма для различных выборок можно ознакомиться в табл. 5.

Также было проведено сравнение разработанного алгоритма классификации с альтернативными подходами: k ближайших соседей, метод опорных векторов, нейронные сети. Данные представлены на рис. 5.

Таблица 4

Результаты работы алгоритма классификации данных с применением различных методов

Данные	Метод	Кол-во термов	Лучшая точность
Iris	DBSCAN	2	0,837
	k-средних	3	0,847
	Mean Shift	2	0,837
Australian	DBSCAN	2	0,783
	k-средних	3	0,779
	Mean Shift	5	0,774
Ionosphere	DBSCAN	3	0,590
	k-средних	4	0,575
	Mean Shift	2	0,792
Seeds	DBSCAN	2	0,558
	k-средних	3	0,647
	Mean Shift	3	0,647
Glass	DBSCAN	3	0,649
	k-средних	3	0,649
	Mean Shift	3	0,649

Таблица 5

Лучшие конфигурации работы алгоритма классификации данных на всех выборках

Данные	Iris		Australian		Ionosphere		Seeds		Glass	
Кол-во термов	4		2		2		4		4	
Кол-во правил	4	5	5	10	4	11	6	7	7	8
Кол-во поколений										
300	0,860	0,863	0,776	0,781	0,787	0,786	0,793	0,805	0,702	0,727
400	0,860	0,863	0,776	0,781	0,787	0,785	0,795	0,805	0,702	0,729
500	0,860	0,863	0,776	0,783	0,790	0,786	0,805	0,805	0,704	0,729
600	0,860	0,863	0,774	0,783	0,790	0,787	0,807	0,805	0,704	0,729
700	0,860	0,863	0,774	0,783	0,792	0,787	0,805	0,805	0,709	0,729
800	0,860	0,863	0,774	0,783	0,792	0,787	0,805	0,805	0,709	0,729

Анализируя данные, представленные на рис. 5, можно сделать вывод, что предлагаемый разработанный алгоритм уступает по точности методам MLP и k -ближайших соседей, но на некоторых наборах данных лучше метода SVM или сравним с ним. Но при увеличении объема ресурсов как минимум до 300 индивидов алгоритм показывает более высокие результаты.

Применение метода Mean Shift уступает в точности разработанному алгоритму, но автоматизирует определение количества термов для различных выборок.

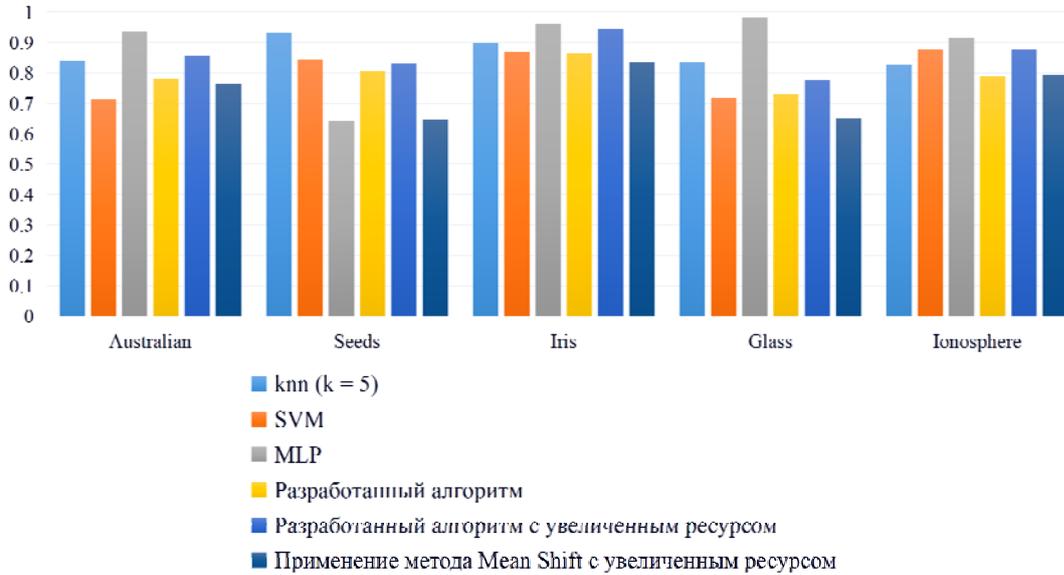


Рис. 5. Визуализация результатов сравнения методов кластеризации

Fig. 5. Visualization of results of comparison of clustering methods

Заключение

В работе предложена алгоритмическая схема классификации данных с применением генетической нечеткой системы и автоматическим определением нечеткой грануляции методом кластеризации данных. Алгоритм протестирован на различных наборах данных с различными параметрами, проведен анализ лучших конфигураций. На основании имеющейся информации была проведена автоматизация определения грануляции разными методами. Сравнение показало, что метод Mean Shift является наиболее подходящим для определения количества термов. Применение данного метода показано небольшое снижение точности на некоторых данных, но его применение позволяет существенно сократить время работы основного алгоритма за счет отсутствия необходимости многократного запуска. Также было проведено сравнение с наиболее популярными алгоритмами для классификации данных – нейронные сети, k ближайших соседей и метод опорных векторов, которое показало, что разработанный алгоритм не сильно уступает в точности, но в будущих работах будет сделан упор на повышение точности классификации. Главное преимущество – возможность интерпретации процесса принятия решений, следовательно, гибридизация эволюционных алгоритмов и нечеткой логики требуют продолжения исследований. В будущих работах также планируется осуществить реализацию альтернативных подходов гибридизации для повышения точности работы алгоритма.

Библиографические ссылки

1. Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd ed, Springer-Verlag, 2009, 764 p.
2. Ashby W. R. An Introduction to Cybernetics. London: Chapman & Hall, 1956, 306 c.
3. Усков А. А., Кузьмин А. В. Интеллектуальные технологии управления. Искусственные нейронные сети и нечеткая логика. М. : Горячая Линия – Телеком, 2004, 143 с.
4. Zadeh L. A. Fuzzy sets // Information and Control. 1965, No. 8 (3). P. 338–353.

5. Рутковская Д., Пилинский М., Рутковский Л. Нейронные сети, генетические алгоритмы и нечеткие системы. М. : Горячая линия – Телеком, 2004, 384 с.
6. Становов В. В. Самонастраивающиеся эволюционные алгоритмы формирования систем на нечеткой логике : дис. ... канд. техн. наук. Красноярск, 2016. 148 с.
7. Гладков Л. А., Курейчик В. В., Курейчик В. М. Генетические алгоритмы. М. : Физматлит, 2006. 368 с.
8. Ishibuchi H., Nojima Y. Analysis of interpretability-accuracy tradeoff of fuzzy systems by multiobjective fuzzy genetics-based machine learning // *International Journal of Approximate Reasoning*. 2007.28 p. Doi:10.1016/j.ijar.2006.01.004.
9. Емельянов В. В., Курейчик В. В., Курейчик В. М. Теория и практика эволюционного моделирования. М. : Физматлит, 2003. 432 с.
10. Генетический алгоритм в оптимизации трехмерной упаковки блоков в контейнер / О. П. Тимофеева, Т. Ю. Чернышева, О. Н. Корелин и др. Нижний Новгород : Нижегородский гос. техн. ун-т им. П. Е. Алексеева, 2017. 7 с.
11. Кочкин А. М. Применение генетического алгоритма в задачах оптимизации. Ч. 2. Караганда : Карагандинский гос. техн. ун-т, 2018. 126 с.
12. Asuncion A., Newman D. UCI machine learning repository. University of California, Irvine, School of Information and Computer Sciences [Электронный ресурс]. 2007. URL: <http://www.ics.uci.edu/~mlern/MLRepository.html>.
13. Плешкова Т. С., Становов В. В. Автоматическое определение грануляции для генетической нечеткой системы с использованием DBSCAN // *Актуальные проблемы авиации и космонавтики : материалы VI Междунар. науч.-практ. конф. творческой молодежи*. Красноярск, 2021. 3 с.
14. Ketchen Jr D. J.; Shook C. L. The application of cluster analysis in Strategic Management Research: An analysis and critique // *Strategic Management Journal*. 1996. No. 17 (6). P. 441–458. Doi: [https://doi.org/10.1002/\(SICI\)1097-0266\(199606\)17:6<441::AID-SMJ819>3.0.CO;2-G](https://doi.org/10.1002/(SICI)1097-0266(199606)17:6<441::AID-SMJ819>3.0.CO;2-G).
15. Gorban A. N., Zinovyev A. Y. Principal Graphs and Manifolds, Ch. 2 in: *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques*, Emilio Soria Olivas et al. IGI Global, Hershey, PA, USA, 2009. 34 p.
16. Cheng Yizong Mean Shift, Mode Seeking, and Clustering // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1995. P. 790–799. Doi: 10.1109/34.400568.

References

1. Hastie T., Tibshirani R., Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer–Verlag, 2009, 764 p.
2. Ashby W. R. *An Introduction to Cybernetics*. London: Chapman & Hall, 1956, 306 p.
3. Uskov A. A., Kuzmin A. V. *Intellectual'nye tekhnologii upravleniya. Iskusstvennye nejronnye seti i nechetkaya logika* [Intellectual management technologies. Artificial neural networks and fuzzy logic]. Moscow, Hotline, Telecom Publ., 2004, 143 p. (In Russ.).
4. Zadeh L. A. Fuzzy sets. *Information and Control*. 1965, No. 8 (3), P. 338–353.
5. Rutkovskaya D., Pilinsky M., Rutkovsky L. *Nejronnye seti, geneticheskie algoritmy i nechetkie sistemy* [Neural networks, genetic algorithms and fuzzy systems]. Moscow, Hotline-Telecom Publ., 2004, 384 p. (In Russ.).
6. Stanovov V. V. *Samonastraivayuschie evolyucionnie algoritmi formirovaniya sistem na nechetkoi logike. Kand. Dis.* [Self-adjusting evolutionary algorithms for the formation of systems based on fuzzy logic. Cand. Diss.]. Krasnoyarsk, 2016, 148 p. (In Russ.).
7. Gladkov L. A., Kureichik V. V., Kureichik V. M. *Geneticheskie algoritmy* [Genetic algorithms. Tutorial. 2nd ed.]. Moscow, Fizmatlit Publ., 2006, 368 p. (In Russ.).

8. Ishibuchi H., Nojima Y. Analysis of interpretability-accuracy tradeoff of fuzzy systems by multiobjective fuzzy genetics-based machine learning. *International Journal of Approximate Reasoning*. 2007, 28 p. Doi: 10.1016/j.ijar.2006.01.004.
9. Emelyanov V. V., Kureychik V. V., Kureychik V. M. *Teoriya i praktika evolyucionnogo modelirovaniya* [Theory and practice of evolutionary modeling]. Moscow, Fizmatlit Publ., 2003, 432 p. (In Russ.).
10. Timofeeva O. P., Chernysheva T. Yu., Korelin O. N., Volkov A. V. *Geneticheskij algoritm v optimizacii trekhmernoj upakovki blokov v kontejner* [A genetic algorithm in optimization of three-dimensional packing of blocks in a container]. Nizhny Novgorod, Nizhny Novgorod state technical university of R. E. Alekseev Publ., 2017, 7 p.
11. Kochkin A. M. *Primenenie geneticheskogo algoritma v zadachakh optimizacii* [Application of the genetic algorithm in optimization problems. Part 2] Karaganda, Karaganda State Technical University Publ., 2018, 126 p.
12. Asuncion A., Newman D. UCI machine learning repository. University of California, Irvine, School of Information and Computer Sciences, 2007. Available at: <http://www.ics.uci.edu/~mllearn/MLRepository.html> (accessed: 15.10.2021).
13. Pleshkova T. S., Stanovov V. V. *Avtomaticeskoe opredelenie granulyacii dlya geneticheskoi nechetkoi sistemi s ispolzovaniem DBSCAN* [Automatic granulation detection for genetic fuzzy system using DBSCAN]. *Materials of the VI international scientific and practical conf. Krasnoyarsk*, 2021, 3 p. (In Russ.).
14. Ketchen Jr D. J.; Shook C. L. The application of cluster analysis in Strategic Management Research: An analysis and critique. *Strategic Management Journal*. 1996, No. 17 (6), P. 441–458. Doi: [https://doi.org/10.1002/\(SICI\)1097-0266\(199606\)17:6<441::AID-SMJ819>3.0.CO;2-G](https://doi.org/10.1002/(SICI)1097-0266(199606)17:6<441::AID-SMJ819>3.0.CO;2-G).
15. Gorban A. N., Zinovyev A. Y. Principal Graphs and Manifolds, Ch. 2 in: *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques*, Emilio Soria Olivas et al. IGI Global, Hershey, PA, USA, 2009, 34 p.
16. Cheng Yizong Mean Shift, Mode Seeking, and Clustering, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1995, 790–799 p. Doi: 10.1109/34.400568.

© Плешкова Т. С., Становов В. В., 2022

Плешкова Татьяна Сергеевна – магистрант; Сибирский федеральный университет. E-mail: tatyana.pleshkova2310@gmail.com.

Становов Владимир Вадимович – кандидат технических наук, доцент базовой кафедры интеллектуальных систем управления; Сибирский федеральный университет. E-mail: vladimirstanovov@yandex.ru.

Pleshkova Tatyana Sergeevna – master student; Siberian Federal University. E-mail: tatyana.pleshkova2310@gmail.com.

Stanovov Vladimir Vadimovich – Ph. D., associate professor, at the base department of intelligent control systems; Siberian Federal University. E-mail: vladimirstanovov@yandex.ru.
