

Синтез базы нечетких правил на основе результатов нечеткой кластеризации регулярных выражений с использованием алгоритмов дифференциальной эволюции

Н.А. Морошкин

Аннотация — Рассматривается задача синтеза базы нечетких правил принадлежности векторных представлений абстрактных синтаксических деревьев регулярных выражений к кластерам на основе результатов нечеткой кластеризации стандартным алгоритмом нечетких С-средних и его модификациями. Основным объектом исследования являются базы нечетких правил на основе алгоритмов Мамдани и Сугено, параметры которых подбираются с использованием эволюционных алгоритмов, включая классический алгоритм дифференциальной эволюции (Differential Evolution, DE) и его модификации L-SRTDE и L-SHADE-RSP. Цель исследования заключается в оценке эффективности применения алгоритмов дифференциальной эволюции для подбора параметров базы нечетких правил с учётом числа кластеров и структурных особенностей регулярных выражений. В исследовании выполнен сравнительный анализ результатов синтеза баз нечетких правил для различных вариантов алгоритмов оптимизации. Качество синтеза оценено с использованием такой метрики как F1-score. Результаты экспериментальных исследований подтверждают целесообразность применения предложенных алгоритмов дифференциальной эволюции для подбора параметров базы нечетких правил, обеспечивающих высокую точность оценки принадлежности регулярных выражений к кластерам.

Ключевые слова — Регулярные выражения, нечеткая кластеризация, база нечетких правил, алгоритм Мамдани, алгоритм Сугено, эволюционный алгоритм, DE, L-SRTDE, L-SHADE-RSP.

1. ВВЕДЕНИЕ

В задачах анализа регулярных выражений (РВ) всё большую актуальность приобретает использование их векторных представлений, формируемых на основе структурных и семантических характеристик абстрактных синтаксических деревьев. Такие представления позволяют применять методы интеллектуального анализа данных, в частности нечеткую кластеризацию, для выявления групп сходных шаблонов. Однако результаты нечеткой кластеризации, особенно при использовании нейросетевых векторных представлений и алгоритмов нелинейного снижения размерности, применяемых для получения векторных

представлений РВ, как правило, обладают низкой интерпретируемостью [1, 2].

Одним из перспективных направлений повышения интерпретируемости векторных представлений [2] РВ является переход от четкого численного описания кластерной структуры (например, полученной алгоритмом К-средних) к лингвистически интерпретируемым моделям, основанным на аппарате нечеткой логики. В этом контексте кластеры могут рассматриваться как нечеткие концепты, а за задачей кластеризации естественным образом следует задача синтеза базы нечетких правил, описывающей принадлежность объекта (в частности, регулярного выражения) к тому или иному кластеру. При этом при решении задачи кластеризации регулярных выражений сначала формируются их векторные представления на основе нейросетевых моделей, а затем к этим векторным представлениям применяются алгоритмы нелинейного снижения размерности, позволяющие получить компактное признаковое пространство.

Целесообразность применения нечеткого моделирования [3] в контексте интерпретации результатов кластеризации регулярных выражений можно описать следующими аспектами.

1. Неясность и «размытость» полученных кластеров. Так, например, регулярные выражения могут обладать высокой степенью структурного сходства, но иметь противоположное семантическое значение.

2. Неполнота структурных и семантических свойств векторных представлений регулярных выражений вследствие работы алгоритмов снижения размерностей.

3. Ограничения структурной интерпретации генератора абстрактных синтаксических деревьев регулярных выражений.

Целью настоящей работы является разработка и исследование метода синтеза базы нечетких правил принадлежности объектов к нечетким кластерам, полученным при нечеткой кластеризации векторных представлений регулярных выражений. В качестве формализма используются базы нечетких правил (БНП) на основе алгоритмов Мамдани и Сугено, а подбор гиперпараметров БНП и функций принадлежности

осуществляется с применением эволюционных алгоритмов на основе дифференциальной эволюции, таких как DE (Differential Evolution) с классической схемой реализации 1/rand/bin, L-SRTDE (Linear population size reduction, Success RaTe-based DE) и L-SHADE-RSP (Success- History based Adaptive DE with Rank-based Selective Pressure). Такой подход позволяет объединить структурную информацию о регулярных выражениях, результаты нечеткой кластеризации и эволюционную оптимизацию в рамках единой интерпретируемой нечеткой модели.

II. БАЗА НЕЧЕТКИХ ПРАВИЛ ПРИНАДЛЕЖНОСТИ ОБЪЕКТА К НЕЧЕТКОМУ КЛАСТЕРУ

Пусть $\vec{x} \in R^d$ ($i = \overline{1, N}$) – векторное представление i -го РВ в некотором d -мерном пространстве, полученное на основе нейросетевого векторного представления абстрактного синтаксического дерева РВ с применением к нему алгоритма UMAP [4] нелинейного снижения размерности для получения компактного низкоразмерного представления, отражающего структуру исходного пространства. В результате применения алгоритма нечеткой кластеризации каждому i -му РВ ставится в соответствие степени принадлежности u_{ij} ($u_{ij} \in [0, 1]; i = \overline{1, N}; j = \overline{1, C}$, где C – число нечетких кластеров, N – число РВ в наборе данных). При этом для любого i -го объекта ($i = \overline{1, N}$) выполняется условие: $\sum_{j=1}^C u_{ij} = 1$.

В рамках данной работы каждый кластер интерпретируется как некоторое нечеткое множество T_j в пространстве признаков (на универсуме R^d), а задача определения степени нечеткой принадлежности u_{ij} i -го объекта к j -му нечеткому множеству T_j формулируется как задача нечеткого вывода. Для решения задачи нечеткого вывода требуется сформировать нечеткий предикат $F: R^d \rightarrow R^C$, где отображение F реализуется в виде системы нечеткого вывода, заданной базой нечетких «если-то» правил (БНП).

Нечёткое правило вида «если-то» задаётся антецедентом (условием), консеквентом (заключением), а также логическими операторами И/ИЛИ, используемыми для связывания атомарных условий. Нечёткое высказывание, описывающее условие или заключение, имеет вид « β есть α », где β – лингвистическая переменная, а α – из её термов. Лингвистическая переменная определяется на некотором универсуме значений и описывается набором термов, каждый из которых представляет собой нечёткое множество, заданное функцией принадлежности. Таким образом, каждой входной и выходной лингвистической переменной соответствует набор нечётких множеств, интерпретируемых как лингвистические категории её значений.

Терм A для переменной x , заданной на универсуме X , можно описать функцией принадлежности:

$$\mu_A(x): X \rightarrow [0, 1], \quad (1)$$

где $\mu_A(x)$ – функция принадлежности, характеризующая степень принадлежности конкретного значения переменной x к терму A .

Наиболее распространенные функции

принадлежности приведены в таблице 1 [5].

Таблица 1 – Перечень функций принадлежности

Название	Параметры	Функция принадлежности $\mu_A(x)$
Гауссова функция	c – центр, σ – ширина	$\mu_A(x) = \exp\left(-\frac{(x-c)^2}{2\sigma^2}\right)$
Треугольная функция	a – левая граница, b – центр, c – правая граница	$\mu_A(x) = \begin{cases} 0, & \text{если } x \leq a \text{ или } x \geq c \\ \frac{x-a}{b-a}, & \text{если } a < x \leq b \\ \frac{c-x}{c-b}, & \text{если } b < x \leq c \end{cases}$
Трапециевидная функция	a – левая нижняя точка, b – левая верхняя точка, c – правая верхняя точка, d – правая нижняя точка	$\mu_A(x) = \begin{cases} 0, & \text{если } x \leq a \text{ или } x \geq d \\ \frac{x-a}{b-a}, & \text{если } a < x \leq b \\ 1, & \text{если } b < x \leq c \\ \frac{d-x}{d-c}, & \text{если } c < x \leq d \end{cases}$
Колоколообразная функция	c – центр, a – ширина, b – крутизна	$\mu_A(x) = \frac{1}{1 + \left \frac{x-c}{a}\right ^{2b}}$

Обычно при разработке БНП применяются алгоритмы Мамдани и Сугено [3].

При синтезе БНП на основе алгоритма Мамдани консеквент нечёткого правила задаётся выходной лингвистической переменной, определённой набором термов, каждый из которых представляет собой нечёткое множество [6]. При синтезе БНП на основе алгоритма Сугено консеквент нечёткого правила задаётся чёткой функцией от входных переменных (или просто константой) [7].

В БНП, основанной на алгоритме Мамдани, в консеквенте каждого нечёткого правила формируется модифицированное нечёткое множество, соответствующее одному из термов выходной лингвистической переменной. Далее осуществляется агрегация нечётких множеств, относящихся к различным термам одной и той же выходной лингвистической переменной, после чего выполняется процедура дефазификации, в результате которой определяется итоговое чёткое значение выходной переменной.

В БНП, основанной на алгоритме Сугено, консеквент нечёткого правила задаётся чёткой функцией входных переменных. Выход каждого правила представляет собой числовое значение данной функции, а результирующий выход БНП вычисляется в виде взвешенного среднего выходов правил, где в качестве весов используются степени их активации. При этом процедура дефазификации не выполняется, поскольку выходы правил уже являются чёткими значениями.

III. ПОДХОД К ПОДБОРУ ГИПЕРПАРАМЕТРОВ БАЗЫ НЕЧЕТКИХ ПРАВИЛ

Синтез БНП предполагает определение таких параметров, как число нечетких правил на каждый кластер, число признаков входного векторного представления на каждое нечеткое правило, число термов, тип функций принадлежности входных и выходных лингвистических переменных. Совокупность

указанных параметров определяет многомерное пространство параметров, а соответствующая задача оптимизации характеризуется высокой размерностью, нелинейной зависимостью от параметров и наличием большого числа локальных экстремумов.

Для решения задачи поиска параметров БНП предлагается применить эволюционные алгоритмы оптимизации [8].

В общем случае задачу, которую решают эволюционные алгоритмы, можно представить в виде:

$$x^* = \operatorname{argmax}_{x \in R} (Q(x)), \quad (2)$$

где x – вектор, принадлежащий пространству поиска R^z и описывающий индивида; z – размерность вектора x ; R^z – область допустимых значений вектора; $Q(x)$ – целевая функция, определяющая значение показателя качества индивида; x^* – вектор оптимального решения.

В рассматриваемой задаче оптимизации параметров нечеткой базы знаний предлагается использовать в качестве целевой функции $Q(x)$ метрику F₁-score.

Эволюционные алгоритмы позволяют эффективно осуществлять глобальный поиск в сложных ландшафтах целевой функции без необходимости вычисления градиентов, что особенно важно при оптимизации гиперпараметров баз нечетких правил. В данной работе предлагается использовать три алгоритма дифференциальной эволюции – алгоритм DE (Differential Evolution) с классической схемой реализации 1/rand/bin и современные алгоритмы L-SRTDE (Linear population size reduction, Success Rate-based DE) и L-SHADE-RSP (Success-History based Adaptive DE with Rank-based Selective Pressure).

Алгоритм дифференциальной эволюции – стохастический алгоритм глобальной оптимизации, реализующий популяционный эволюционный поиск с применением операторов мутации в пространстве признаков [9]. Алгоритм дифференциальной эволюции выполняет поиск минимума целевой функции в некотором q -мерном пространстве решений.

Алгоритм L-SRTDE является модификацией алгоритма дифференциальной эволюции, в котором основное внимание уделено адаптации масштабирующего коэффициента мутации, отличающегося высокой чувствительностью к настройке [10]. В данном алгоритме значение масштабирующего коэффициента определяется на основе показателя успешности, вычисляемого как отношение числа улучшенных решений к размеру популяции в текущем поколении. Такой подход позволяет динамически настраивать параметры алгоритма в процессе оптимизации без использования истории успешных параметров.

Алгоритм L-SHADE-RSP представляет собой модификацию алгоритма L-SHADE и основан на использовании стратегии мутации с адаптивным уменьшением размера популяции и историей успешных параметров [11]. Ключевой особенностью L-SHADE-RSP является применение рангоориентированного селективного давления (Rank-based Selective Pressure, RSP), которое используется при выборе опорных векторов для генерации новых кандидатов. Вероятность

выбора решений определяется их рангом в текущей популяции, что позволяет более гибко управлять балансом между исследованием и эксплуатацией пространства поиска.

Пусть в нечетком правиле БНП может быть задействовано несколько входных лингвистических переменных, на основе которых определяются подусловия правила, и одна выходная лингвистическая переменная, на основе которой определяется заключение правила.

При использовании алгоритмов дифференциальной эволюции для синтеза БНП необходимо определить структуру хромосомы индивида.

В хромосоме для БНП типа Мамдани кодируются следующие параметры.

1. Число нечетких правил для каждого кластера r_{num}^c ($r_{num}^c \in [r_{min}^c, r_{max}^c]$, где r_{min}^c – минимальное число нечетких правил для каждого кластера; r_{max}^c – максимальное число нечетких правил для каждого кластера). Этот параметр вводится для ограничения необходимых вычислительных ресурсов.

2. Число признаков на одно правило l_{num}^c ($l_{num}^c \in [l_{min}^c, l_{max}^c]$, где l_{min}^c – минимальное число признаков на каждое правило; l_{max}^c – максимальное число признаков на каждое правило). Этот параметр вводится для ограничения необходимых вычислительных ресурсов.

3. Номер используемой функции принадлежности n_{input} ($n_{input} \in \{1, 2, 3, 4, 5\}$) для входного термина, соответствующий номеру функции принадлежности из таблицы 1.

4. Номер используемой функции принадлежности n_{output} ($n_{output} \in \{1, 2, 3, 4, 5\}$) для выходного термина, соответствующий номеру функции принадлежности из таблицы 1.

5. Вероятность оператора связки «ИЛИ» $p_{or} \in [0, 1]$. По значению данного параметра вычисляется вероятность оператора связки «И» $p_{and} = 1 - p_{or}$. Выбор оператора связки зависит от параметра БНП $p_{or} \in [0, 1]$, данный параметр выбирается до процесса синтеза БНП и остается постоянным в процессе синтеза, p_{or} – вероятность оператора связки «ИЛИ». В процессе синтеза нечетких правил и записи подусловия генерируется случайное число $\theta \in [0, 1]$, если $\theta \geq p_{or}$, то выбирается оператор связки «ИЛИ», в ином случае – «И».

6. Булев вектор \vec{t}_y ($y = \overline{1, Y}$), определяющий использование признаков в нечетких правилах для каждого кластера, длиной $Y = r_{max}^c \cdot l_{max}^c \cdot C$, где C – число кластеров. Значение элемента вектора равно 1, если соответствующий признак используется в заданном правиле кластера, и 0 – в противном случае.

В общем виде хромосому для синтеза БНП алгоритма Мамдани можно представить в виде кортежа $(r_{num}^c, l_{num}^c, n_{input}, n_{output}, p_{or}, \vec{t}_y)$.

В данной работе предлагается использовать БНП алгоритма Сугено первого порядка, в этом случае выход каждого правила задается параметрической линейной функцией входных переменных. При этом для k -го

правила используется линейная модель первого порядка, определяющаяся линейной функцией:

$$f_k(\vec{x}) = a_{k_0} + \sum_{j=1}^d a_{kj} x_j \quad (3)$$

где $\vec{x} \in R^d$ – входной вектор размерностью d , a_{k_0} – свободный член, a_{kj} – коэффициент при входной переменной x_j .

Тогда вектор коэффициентов k -го правила имеет вид:

$$\vec{a}_k = (a_{k_0}, \dots, a_{k_d}) \in R^{d+1}. \quad (4)$$

Вектор \vec{a}_k содержит K элементов, где $K = C \cdot (d + 1)$, где C – число кластеров, а d – размерность пространства векторных представлений РВ.

В хромосоме для БНП алгоритма Сугено кодируются следующие параметры.

1. Число нечетких правил для каждого кластера r_{num}^c ($r_{num}^c \in [r_{min}^c, r_{max}^c]$, где r_{min}^c – минимальное число нечетких правил для каждого кластера; r_{max}^c – максимальное число нечетких правил для каждого кластера). Этот параметр вводится для ограничения необходимых вычислительных ресурсов.

2. Число признаков на одно правило l_{num}^c ($l_{num}^c \in [l_{min}^c, l_{max}^c]$, где l_{min}^c – минимальное число признаков на каждое правило; l_{max}^c – максимальное число признаков на каждое правило). Этот параметр вводится для ограничения необходимых вычислительных ресурсов.

3. Номер используемой функции принадлежности n_{input} ($n_{input} \in \{1, 2, 3, 4, 5\}$) для входного термина, соответствующий номеру функции принадлежности из таблицы 1.

4. Вероятность оператора связки «ИЛИ» $p_{or} \in [0, 1]$. По значению данного параметра вычисляется вероятность оператора связки «И» $p_{and} = 1 - p_{or}$. Выбор оператора связки зависит от параметра БНП $p_{or} \in [0, 1]$, данный параметр выбирается до процесса синтеза БНП и остается постоянным в процессе синтеза, p_{or} – вероятность оператора связки «ИЛИ». В процессе синтеза нечетких правил и записи подусловия генерируется случайное число $\theta \in [0, 1]$, если $\theta \geq p_{or}$, то выбирается оператор связки «ИЛИ», в ином случае – «И».

5. Вектора коэффициентов линейной модели выходного термина \vec{a}_k ($a_{kj} \in [0, 1], k = 1, K$).

6. Булев вектор \vec{t}_y ($y = \overline{1, Y}$), определяющий использование признаков в нечетких правилах для каждого кластера, длиной $Y = r_{max}^c \cdot l_{max}^c \cdot C$, где C – число кластеров. Значение элемента вектора равно 1, если соответствующий признак используется в заданном правиле, и 0 – в противном случае.

В общем виде хромосому для синтеза БНП алгоритма Сугено можно представить в виде кортежа $(r_{num}^c, l_{num}^c, n_{input}, p_{or}, \vec{a}_k, \vec{t}_y)$.

В выбранных алгоритмах дифференциальной эволюции (DE, L-SRTDE, L-SHADE-RSP) каждый ген хромосомы кодируется вещественным числом. Так как некоторые параметры, закодированные в генах, являются целыми числами, необходимо использовать функцию преобразования вещественных чисел в целые. Наиболее популярной и простой функцией преобразования для получения дискретных порядковых переменных

является функция усечения (truncation) $y_i = INT(y_i)$, где $INT(\cdot)$ – функция, преобразующая вещественное значение в целое посредством округления до ближайшего доступного значения [12].

IV. ЭКСПЕРИМЕНТАЛЬНЫЕ ИССЛЕДОВАНИЯ

Для проведения экспериментальных исследований был выбран набор данных, полученный с применением модели BERT (с модификацией base-code) [13]. При этом сами регулярные выражения были представлены в виде абстрактных синтаксических деревьев без какой-либо постобработки. Для этого набора данных были получены результаты нечеткой кластеризации алгоритмом нечетких С-средних и его модификаций. При этом были использованы следующие версии алгоритмов нечеткой кластеризации:

1. Стандартный алгоритм нечетких С-средних [14].
2. Алгоритм С-средних с L2-регуляризацией [15].
3. Алгоритм С-средних с использованием энтропии [15].
4. Алгоритм С-средних со штрафом на принадлежность объекта ко многим кластерам [15].
5. Алгоритм С-средних с использованием штрафа на большие размеры кластеров [15].

В качестве метода оптимизации центроидов кластеров использовались подходы с применением градиентных методов (градиентный спуск, Adam, RMSProp), алгоритмы дифференциальной эволюции (DE, L-SRTDE, L-SHADE-RSP) и гибридные подходы. Кластеризация производилась на 2 и 5 кластеров. Всего было получено 160 результатов нечеткой кластеризации. Для экспериментального исследования на основе каждого результата нечеткой кластеризации производилось 10 запусков трёх предложенных к использованию алгоритмов дифференциальной эволюции для подбора гиперпараметров для БНП на основе алгоритма Мамдани и для БНП на основе алгоритма Сугено.

Экспериментальные исследования были выполнены с применением языка программирования C++ версии 17 в среде разработки CLion. В ходе них был использован компьютер со следующими характеристиками: MacBook Air 13 2020 A2337 (процессор: Apple M1 3.2 ГГц 5 нм, ARMv8.5-A, 3.2 ГГц, 8 ядер; оперативная память: 8 Гб; 64-разрядная операционная система). Для реализации БНП на основе алгоритмов Мамдани и Сугено была использована библиотека FuzzyLite [16].

Для алгоритма DE использованы следующие настройки параметров:

1. Коэффициент масштабирования: $F = 0.5$.
2. Вероятность кроссовера: $Cr = 0.8$.

Для алгоритма L-SRTDE настройки параметров аналогичны настройкам алгоритма DE.

Для алгоритма L-SHADE-RSP использованы следующие настройки параметров:

1. Размер архива: $s_{arch} = 1.0$.
2. Вероятность архивации: $p_{arch} = 0.25$.
3. Параметр $p_{size} = 0.17$.

Остальные настройки совпадают с настройками алгоритма L-SRTDE.

Для алгоритмов Мамдани и Сугено для синтеза БНП были использованы следующие границы диапазонов для параметров r_{num}^c и l_{num}^c :

1. Минимальное число нечетких правил для каждого кластера $r_{min}^c = 1$.
2. Максимальное число нечетких правил для каждого кластера $r_{max}^c = 10$.
3. Минимальное число признаков на каждое правило $l_{min}^c = 1$.
4. Максимальное число признаков на каждое правило $l_{max}^c = 10$.

Объекты, полученные после применения алгоритма нелинейного снижения размерности UMAP к векторным представлениям РВ, размерность которых равна 768, имеют размерность d , равную 64. Снижение размерности векторных представлений РВ до 64 позволяет сохранить основную вариативность данных и снизить потребление вычислительных ресурсов.

При этом в новом пространстве размерностью d значения признаков $l_y(y = \overline{1, d})$ всех объектов принадлежат диапазону $[0, 1]$. При синтезе БНП один из подбираемых параметров – число признаков, входящих в состав подусловий в нечетком правиле. Пусть в процессе синтеза БНП минимальное число признаков в нечетком правиле определяется как $r_{min} = 1$, а максимальное число признаков в нечетком правиле определяется как $r_{max} = 10$.

Таким образом в любом нечетком правиле может быть 10 подусловий вида $\alpha = \beta$, где α – название лингвистической переменной, а β – название термина. Пусть число термов T равно 5 при синтезе БНП для 5 кластеров и число термов t равно 2 при для синтеза БНП для 2 кластеров. Параметры функций принадлежности определяются до экспериментов и остаются неизменными во время проведения экспериментальных исследований. При этом тип функций принадлежности выбирается на этапе инициализации синтеза БНП и остается неизменным для всех термов всех лингвистических переменных.

Подусловия могут быть связаны операторами связки «И»/«ИЛИ». Выбор оператора связки зависит от параметра БНП p_{or} ($p_{or} \in [0, 1]$), который выбирается до процесса синтеза БНП и остается постоянным в процессе синтеза. При синтезе каждого нечеткого правила для связывания τ -го и $(\tau + 1)$ -го подусловий ($\tau = \overline{1, r - 1}$, где r – число подусловий) генерируются случайные числа θ_τ ($\theta_\tau \in [0, 1]$). При этом если $\theta_\tau \geq p_{or}$, то выбирается оператор связки «ИЛИ», в ином случае – «И».

Пусть для каждой лингвистической переменной, описывающей некоторый признак $l_y(y = \overline{1, d})$ объекта, в подусловии БНП с 2 кластерами задается два термина: «низкое значение», и «высокое значение», функции принадлежности и значения параметров которых могут быть определены в соответствии с таблицей 2. Пусть для каждой лингвистической переменной, описывающей некоторый признак $l_y(y = \overline{1, d})$ объекта, в подусловии БНП с 5 кластерами задается пять термов: «очень низкое значение», «низкое значение», «среднее значение», «высокое значение» и «очень высокое значение», функции принадлежности (и значения параметров которых могут быть определены в соответствии с таблицей 2. Пусть T – количество термов для входной лингвистической переменной, t – номер термина, тогда центры термов распределяются равномерно и центр t -го термина $c_t = \frac{t-1}{T-1}$, а расстояние между центрами $\Delta = \frac{1}{T-1}$.

В таблице 2 показаны названия используемых в эксперименте функций принадлежности (таблица 1) и значения их параметров с учётом количества термов для входных лингвистических переменных в БНП на основе алгоритмов Мамдани и Сугено.

Таблица 2 – Функции принадлежности с указанием значений их параметров для входных лингвистических переменных для БНП на основе алгоритмов Мамдани и Сугено

Название	Параметры
Гауссова функция	$c = c_t$ – центр; $\sigma = 0.5 \cdot \Delta$ – ширина.
Треугольная функция	$a = c_t - \Delta$ – левая граница; $b = c_t$ – центр; $c = c_t + \Delta$ – правая граница.
Трапециевидная функция	$a = c_t - \Delta$ – левая нижняя точка; $b = c_t - \Delta/4$ – левая верхняя точка; $c = c_t + \Delta/4$ – правая верхняя точка; $d = c_t + \Delta$ – правая нижняя точка.
Колоколообразная функция	$c = c_t$ – центр; $a = 0.7 \cdot \Delta$ – ширина; $b = 2.0$ – крутизна.

Пусть C – количество термов (количество кластеров) во входной лингвистической переменной для БНП на основе алгоритма Мамдани, c – номер термина (номер кластера), тогда центры термов распределяются равномерно, а расстояние между центрами $\Delta = 1$.

В таблице 3 показаны названия используемых в эксперименте функций принадлежности (таблица 1) и значения их параметров с учётом количества термов для выходных лингвистических переменных в БНП на основе алгоритма Мамдани (для БНП на основе алгоритма Сугено выходные лингвистические переменные описываются четкими функциями).

Таблица 3 – Функции принадлежности с указанием значений их параметров для выходных лингвистических переменных для БНП на основе алгоритма Мамдани

Название	Параметры
Гауссова функция	$c = c_t$ – центр; $\sigma = 0.5$ – ширина.
Треугольная функция	$a = c_t - 1$ – левая граница; $b = c_t$ – центр; $c = c_t + 1$ – правая граница.
Колоколообразная функция	$c = c_t$ – центр; $a = 0.7$ – ширина; $b = 2.0$ – крутизна.

Пусть заключение нечеткого правила имеет вид $A = B$, где A – лингвистическая переменная, описывающая РВ, а B – название термина, определяемое как «кластер_с», $c \in \overline{1, C}$, где C – число кластеров.

В БНП, основанной на алгоритме Мамдани, значение принадлежности u_c ($u_c \in [0, 1]$) к c -му кластеру определяется функцией принадлежности, а окончательный числовой выход системы получается после операции классической деффагификации по методу центра тяжести (centroid method) и означает номер нечеткого кластера. В БНП, основанной на алгоритме Сугено, выходная принадлежность к c -му кластеру u_c задаётся линейной функцией входных переменных в случае Сугено 1-го порядка и итоговое четкое значение также означает номер нечеткого кластера. Термами выходных лингвистических переменных являются номера нечетких кластеров.

Так, в частном случае в БНП с 2 кластерами нечеткие правила с 2 подусловиями, определяемыми на основе признаков векторных представлений РВ, могут иметь вид:

1. Если «признак_1» имеет «высокое значение» И «признак_2» имеет «высокое значение», то «РВ» входит в «кластер_1».
2. Если «признак_3» имеет «низкое значение» ИЛИ «признак_4» имеет «низкое значение», то «РВ» входит в «кластер_2».

Конфигурация параметров лучшей (в смысле максимального значения метрики F1-score) синтезированной БНП алгоритма Мамдани для двух кластеров выглядит следующим образом:

1. Число нечетких правил для каждого кластера $r_{num}^C = 1$.
2. Число признаков на одно правило $l_{num}^C = 4$.
3. Номер используемой функции принадлежности для термина входной лингвистической переменной (номер функции принадлежности из таблицы 1) $n_{input} = 3$.
4. Номер используемой функции принадлежности для термина выходной лингвистической переменной (номер функции принадлежности из таблицы 1) $n_{output} = 1$.
5. Вероятность оператора связки ИЛИ $p_{or} = 0$.

Конфигурация параметров лучшей (в смысле максимального значения метрики F1-score) синтезированной БНП алгоритма Мамдани для пяти кластеров выглядит следующим образом:

1. Число нечетких правил для каждого кластера $r_{num}^C = 1$.
2. Число признаков на одно правило $l_{num}^C = 3$.
3. Номер используемой функции принадлежности для термина входной лингвистической переменной (номер функции принадлежности из таблицы 1) $n_{input} = 1$.
4. Номер используемой функции принадлежности для термина выходной лингвистической переменной (номер функции принадлежности из таблицы 1) $n_{output} = 3$.
5. Вероятность оператора связки ИЛИ $p_{or} = 0.5$.

Конфигурация параметров лучшей (в смысле максимального значения метрики F1-score) синтезированной БНП алгоритма Сугено для двух кластеров выглядит следующим образом:

1. Число нечетких правил для каждого кластера $r_{num}^C = 1$.
2. Число признаков на одно правило $l_{num}^C = 1$.
3. Номер используемой функции принадлежности для входного термина (номер функции принадлежности из таблицы 1) $n_{input} = 1$.
4. Вероятность оператора связки ИЛИ $p_{or} = 0$.

5. Вектор коэффициентов линейной модели \vec{a}_k выходной переменной длиной $K = 130$ визуализирован на рисунке 1, на котором показано распределение коэффициентов вектора \vec{a}_k по индексам.

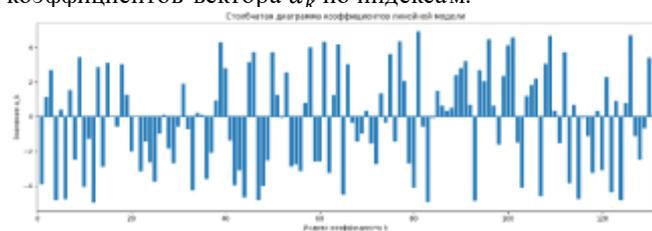


Рисунок 1 – Столбчатая диаграмма коэффициентов линейной модели для выходной переменной при синтезе БНП на основе алгоритма Сугено для двух кластеров.

Конфигурация параметров лучшей (в смысле максимального значения метрики F1-score) синтезированной БНП алгоритма Сугено для пяти кластеров выглядит следующим образом:

1. Число нечетких правил для каждого кластера $r_{num}^C = 1$.
2. Число признаков на одно правило $l_{num}^C = 4$.
3. Номер используемой функции принадлежности для входного термина (номер функции принадлежности из таблицы 1) $n_{input} = 1$.
4. Вероятность оператора связки ИЛИ $p_{or} = 0.4$.
5. Вектор коэффициентов линейной модели \vec{a}_k выходной переменной длиной $K = 325$ визуализирован на рисунке 2, показано распределение коэффициентов вектора \vec{a}_k по индексам.

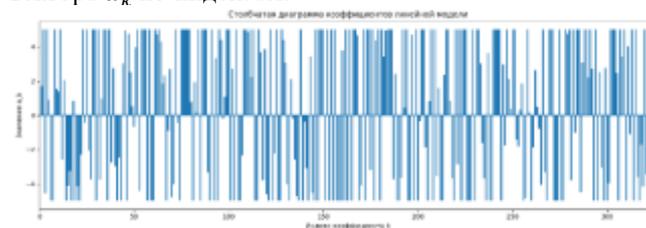


Рисунок 2 – Столбчатая диаграмма коэффициентов линейной модели для выходной переменной при синтезе БНП на основе алгоритма Сугено для пяти кластеров.

В таблицах 4 и 5 показаны лучшие БНП (в смысле обеспечения максимального значения метрики F1-score) на основе алгоритма Мамдани для 2 кластеров и для 5 кластеров соответственно.

Таблица 4 – Синтезированная БНП на основе алгоритма Мамдани для 2 кластеров

№ правила	Нечеткое правило
1	Если «признак_2» имеет «низкое значение» И «признак_4» имеет «высокое значение» И «признак_6» имеет «низкое значение» и «признак_17» имеет «низкое значение», то «РВ» входит в «кластер_1»
2	Если «признак_2» имеет «высокое значение» И «признак_3» имеет «низкое значение» И «признак_4» имеет «высокое значение» И «признак_17» имеет «высокое значение», то «РВ» входит в «кластер_1»

Таблица 5 – Синтезированная БНП на основе алгоритма Мамдани для 5 кластеров

№ правила	Нечеткое правило
1	Если «признак_1» имеет «высокое значение» И «признак_3» имеет «высокое значение» ИЛИ «признак_7» имеет «низкое значение», то «РВ» входит в «кластер_1»
2	Если «признак_1» имеет «высокое значение» И «признак_3» имеет «среднее значение» ИЛИ «признак_5» имеет «низкое значение», то «РВ» входит в «кластер_2»
3	Если «признак_1» имеет «низкое значение» И «признак_3» имеет «высокое значение» ИЛИ «признак_8» имеет «высокое значение», то «РВ» входит в «кластер_3»
4	Если «признак_1» имеет «среднее значение» И «признак_3» имеет «среднее значение» ИЛИ «признак_7» имеет «высокое значение», то «РВ» входит в «кластер_4»
5	Если «признак_1» имеет «высокое значение» И «признак_2» имеет «высокое значение» ИЛИ

	«признак_8» имеет «среднее значение», то «РВ» входит в «кластер_5»
--	--

В таблицах 6 и 7 представлены лучшие варианты БНП (в смысле обеспечения максимального значения метрики F_1 -score), сформированные на основе алгоритма Сугено при числе нечетких кластеров, равном 2 и 5 соответственно, и обеспечивающие максимальное значение метрики F_1 -score на обучающей выборке.

Таблица 6 – Синтезированная БНП на основе алгоритма Сугено для 2 кластеров

№ правила	Нечеткое правило
1	Если «признак_6» имеет «низкое значение», то «РВ» входит в «кластер_1»
2	Если «признак_6» имеет «высокое значение», то «РВ» входит в «кластер_2»

Таблица 7 – Синтезированная БНП на основе алгоритма Сугено для 5 кластеров

№ правила	Нечеткое правило
1	Если «признак_1» имеет «высокое значение» ИЛИ «признак_2» имеет «низкое значение» И «признак_4» имеет «низкое значение» ИЛИ «признак_9» имеет «высокое значение», то «РВ» входит в «кластер_2»
2	Если «признак_1» имеет «высокое значение» ИЛИ «признак_2» имеет «низкое значение» И «признак_4» имеет «среднее значение» И «признак_15» имеет «очень высокое значение», то «РВ» входит в «кластер_2»
3	Если «признак_1» имеет «высокое значение» ИЛИ «признак_2» имеет «высокое значение» И «признак_4» имеет «очень низкое значение» И «признак_15» имеет «низкое значение», то «РВ» входит в «кластер_2»
4	Если «признак_6» имеет «высокое значение» ИЛИ «признак_17» имеет «среднее значение» И «признак_10» имеет «среднее значение» И «признак_18» имеет «среднее значение», то «РВ» входит в «кластер_2»
5	Если «признак_6» имеет «высокое значение» ИЛИ «признак_17» имеет «высокое значение» И «признак_10» имеет «высокое значение» И «признак_18» имеет «очень высокое значение», то «РВ» входит в «кластер_2»

На рисунках 4 и 5 представлены графические зависимости для значений метрики F_1 -score по поколениям для 10 наилучших подходов, полученных при синтезе БНП на основе алгоритма Мамдани в случае кластеризации на 2 кластера и 5 кластеров соответственно. Значения для числа нечетких кластеров, равные 2 и 5, выбраны исходя из необходимости анализа контрастных по сложности структур БНП, а также с учётом вычислительных ограничений, связанных с процедурой синтеза и обучения нечеткой модели. На рисунках 6 и 7 представлены графические зависимости для значений метрики F_1 -score по поколениям для 10 наилучших подходов, полученных при синтезе БНП на основе алгоритма Сугено при кластеризации на 2 кластера и 5 кластеров соответственно. Перечень сокращений, используемых на рисунках 1–4, приведён в таблице 8.

Для БНП на основе алгоритма Мамдани видно, что в случае кластеризации на 2 кластера достигаются более высокие значения метрики F_1 -score, чем в случае кластеризации на 5 кластеров. На рисунке 6 наблюдается устойчивый рост значений метрики F_1 -score по мере

роста числа итераций, причём для нескольких лучших конфигураций параметров алгоритма Мамдани метрика F_1 -score достигает значений около 0.91.

Таблица 8 – Перечень используемых сокращений на рисунках 3, 4, 5 и 6.

Сокращенное название	Алгоритм кластеризации	Градиентный метод оптимизации центроидов
fcm	Стандартный алгоритм С-средних	-
fcml2reg	Алгоритм С-средних с L2-регуляризацией	
fcmentropy	Алгоритм С-средних с использованием энтропии	
fcm sparsity	Алгоритм С-средних со штрафом на принадлежность объекта ко многим кластерам	
fcmlarge	Алгоритм С-средних с использованием штрафа на большие размеры кластеров	
fcm_gd	Стандартный алгоритм С-средних	Метод GD
fcml2reg_gd	Алгоритм С-средних с L2-регуляризацией	
fcmentropy_gd	Алгоритм С-средних с использованием энтропии	
fcm sparsity_gd	Алгоритм С-средних со штрафом на принадлежность объекта ко многим кластерам	
fcmlarge_gd	Алгоритм С-средних с использованием штрафа на большие размеры кластеров	
fcm_adam	Стандартный алгоритм С-средних	Метод Adam
fcml2reg_adam	Алгоритм С-средних с L2-регуляризацией	
fcmentropy_adam	Алгоритм С-средних с использованием энтропии	
fcm sparsity_adam	Алгоритм С-средних со штрафом на принадлежность объекта ко многим кластерам	
fcmlarge_adam	Алгоритм С-средних с использованием штрафа на большие размеры кластеров	
fcm_rmsprop	Стандартный алгоритм С-средних	Метод RMSProp
fcml2reg_rmsprop	Алгоритм С-средних с L2-регуляризацией	
fcmentropy_rmsprop	Алгоритм С-средних с использованием энтропии	
fcm sparsity_rmsprop	Алгоритм С-средних со штрафом на принадлежность объекта ко многим кластерам	
fcmlarge_rmsprop	Алгоритм С-средних с использованием штрафа на большие размеры кластеров	

При этом для 5 кластеров (рисунок 6) максимальные значения метрики F_1 -score не превышают 0.75, а рост значений метрики F_1 -score по итерациям менее выражен. Это свидетельствует о том, что при большем числе кластеров повышается сложность синтеза правил, что затрудняет достижение высокой точности. Из рисунков 4 и 5 видно, что модификации алгоритма дифференциальной эволюции (LSHADE-RSP, L-SRTDE) по сравнению с классическим алгоритмом DE обеспечивают более высокие значения метрики F_1 -score для БНП на основе алгоритма Мамдани.

Для БНП на основе алгоритма Сугено также наблюдается существенное влияние числа кластеров на качество синтеза. В случае кластеризации на 2 кластера достигаются значительно более высокие значения

метрики F_1 -score по сравнению с вариантом с кластеризацией на 5 кластеров.

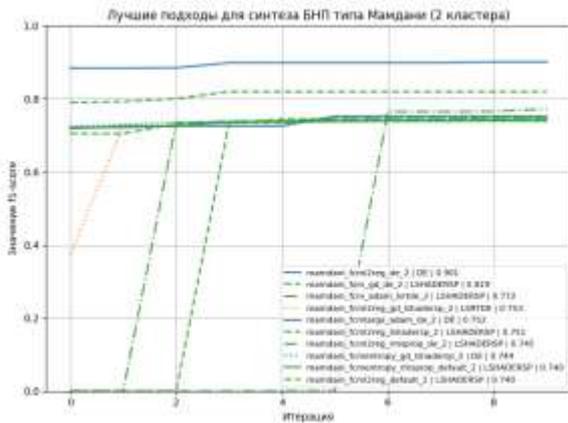


Рисунок 3 – Графические зависимости для значений метрики F_1 -score по поколениям для 10 наилучших подходов, полученных при синтезе БНП на основе алгоритма Мамдани для результатов нечеткой кластеризации на 2 кластера.

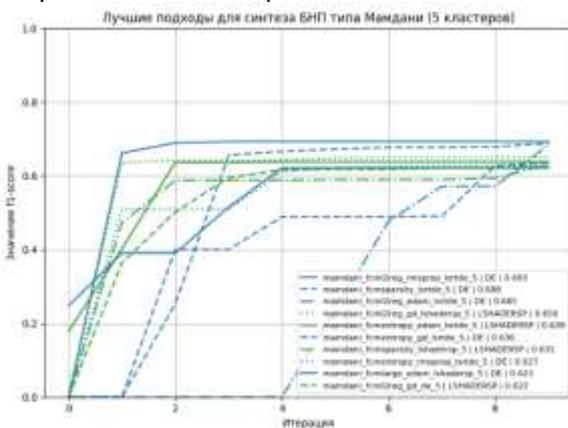


Рисунок 4 – Графические зависимости для значений метрики F_1 -score по поколениям для 10 наилучших подходов, полученных при синтезе БНП на основе алгоритма Мамдани для результатов нечеткой кластеризации на 5 кластера.

На рисунке 6 видно устойчивое увеличение значения метрики F_1 -score по мере роста числа итераций, причём для лучших подходов значения F_1 -score достигают уровня около 0.88–0.89.

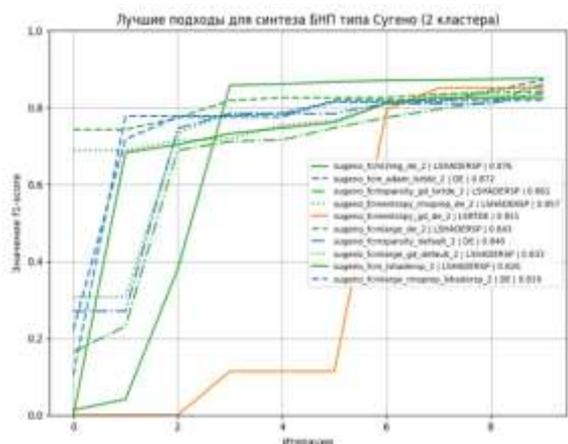


Рисунок 5 – Графические зависимости для значений метрики F_1 -score по поколениям для 10 наилучших подходов, полученных при синтезе БНП на основе

алгоритма Сугено для результатов нечеткой кластеризации на 2 кластера.

При этом рост значений метрики F_1 -score носит выраженный и стабильный характер уже на ранних итерациях.

В случае использования 5 кластеров (рисунок 6) максимальные значения метрики F_1 -score заметно ниже и не превышают примерно 0.66, а динамика роста значений метрики F_1 -score по итерациям менее выражена. Это указывает на то, что увеличение числа кластеров приводит к усложнению структуры правил Сугено и ухудшению обобщающей способности, что затрудняет достижение высокой точности классификации.

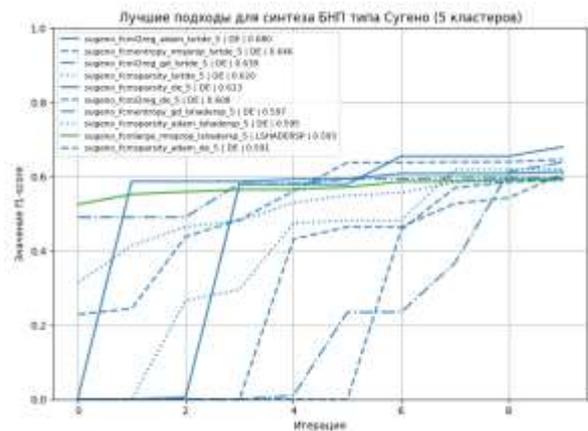


Рисунок 6 – Графические зависимости для значений метрики F_1 -score по поколениям для 10 наилучших подходов, полученных при синтезе БНП на основе алгоритма Сугено для результатов нечеткой кластеризации на 5 кластеров.

Таким образом, результаты экспериментов показывают, что число кластеров является одним из ключевых факторов, определяющих качество синтеза баз нечетких правил как для алгоритма Мамдани, так и для алгоритма Сугено. В обоих случаях кластеризация на 2 кластера обеспечивает существенно более высокие значения метрики F_1 -score и более устойчивую положительную динамику её роста по итерациям по сравнению с вариантом кластеризации на 5 кластеров. Увеличение числа кластеров приводит к усложнению структуры правил, росту размерности пространства параметров и, как следствие, снижению обобщающей способности модели, что ограничивает достижимый уровень точности классификации.

V. ЗАКЛЮЧЕНИЕ

В работе исследована задача синтеза баз нечетких правил для оценки принадлежности векторных представлений абстрактных синтаксических деревьев регулярных выражений к кластерам, полученным методами нечеткой кластеризации. Проведённый сравнительный анализ показал, что эффективность синтеза БНП существенно зависит как от типа нечеткого вывода (Мамдани, Сугено), так и от числа кластеров, используемых на этапе кластеризации. Экспериментально установлено, что кластеризация на 2 кластера обеспечивает более высокие значения метрики F_1 -score и более устойчивую сходимость по итерациям по сравнению с вариантом кластеризации на 5 кластеров, для которого характерно

усложнение структуры правил и снижение обобщающей способности.

Кроме того, показано, что применение современных модификаций алгоритма дифференциальной эволюции (L-SRTDE и L-SHADE-RSP) для подбора параметров БНП является более эффективным по сравнению с применением классического алгоритма ДЭ. Указанные алгоритмы демонстрируют лучшие итоговые значения метрики F₁-score и более стабильную динамику оптимизации как для БНП на основе алгоритма Мамдани, так и для БНП на основе алгоритма Сугено. Полученные результаты подтверждают целесообразность использования эволюционных алгоритмов оптимизации при синтезе БНП и могут быть использованы при построении интеллектуальных систем анализа и классификации структурно сложных объектов, в частности – регулярных выражений.

БИБЛИОГРАФИЯ

- [1] Л. А. Демидова, В. В. Кираковский и А. Н. Пылькин, Принятие решений в условиях неопределенности, 2-е изд., перераб. Москва, 2015.
- [2] К. А. Никитенко и А. В. Звягинцева, “Интерпретируемость нейросемантических моделей при их применении в прикладных областях,” Проблемы искусственного интеллекта, № 2(37), с. 79–90, 2025.
- [3] L. A. Demidova, Yu. S. Sokolova, and E. V. Nikulchev, “Use of fuzzy clustering algorithms ensemble for SVM classifier development,” International Review on Modelling and Simulations, vol. 8, no. 4, pp. 446–457, 2015.
- [4] L. A. Demidova and A. V. Gorchakov, “Fuzzy information discrimination measures and their application to low dimensional embedding construction in the UMAP algorithm,” Journal of Imaging vol. 8, p. 113, 2022.
- [5] А. В. Леоненков, Нечеткое моделирование в среде MATLAB и fuzzyTECH. СПб.: БХВ-Петербург, 2005, 736 с.
- [6] Н. С. Дыбулина и С. А. Царева, “Модель оценки результативности процессов системы менеджмента качества на основе применения алгоритма Мамдани в среде MATLAB,” Вестник Магнитогорского государственного технического университета им. Г. И. Носова, т. 23, № 2, с. 148–157, 2025.
- [7] М. Х. Чан, Т. Ф. Ку и Т. Т. Нгуен, “Применение алгоритма нечеткого логического вывода Сугено в задаче прогнозирования остаточного ресурса авиационных радиотехнических средств,” Надежность и качество сложных систем, № 2(50), с. 15–25, 2025.
- [8] А. Ф. Антипин, Е. В. Антипина и С. А. Мустафина, “Модифицированный нечеткий регулятор с оптимизацией режимных параметров технологического процесса,” Мехатроника, автоматизация, управление, т. 25, № 7, с. 354–361, 2024.
- [9] К. Е. Долженко, А. В. Кожина, Д. Р. Новиков и В. Р. Козлов, “Алгоритм дифференциальной эволюции для решения задач машинного обучения,” Научно-технический вестник Поволжья, № 7, с. 116–118, 2025.
- [10] V. Stanovov and E. Semenkin, “Success rate-based adaptive differential evolution L-SRTDE for CEC 2024 competition,” in Proc. 2024 IEEE Congress on Evolutionary Computation (CEC), Yokohama, Japan, Jun. 30–Jul. 5, 2024, pp. 1–8.
- [11] V. Stanovov, S. Akhmedova, and E. Semenkin, “NL-SHADE-RSP algorithm with adaptive archive and selective pressure for CEC 2021 numerical optimization,” in Proc. 2021 IEEE Congress on Evolutionary Computation (CEC), Kraków, Poland, Jun. 28–Jul. 1, 2021, pp. 809–816.
- [12] J. Lampinen and I. Zelinka, “Mixed integer-discrete-continuous optimization by differential evolution,” in Proc. 5th Int. Conf. on Soft Computing, 1999.
- [13] Л. А. Демидова и Н. А. Морошкин, “Решение задачи кластеризации векторных представлений регулярных выражений,” Вестник Воронежского государственного технического университета, № 4(21), с. 50–59, 2025.
- [14] Т. Т. З. Нгуен и Л. В. Черненькая, “Модель анализа факторов на основе нечеткой кластеризации С-средних,” Известия Тульского государственного университета, № 1, с. 329–337, 2023.
- [15] E. Bedalli, S. Hajrulla, R. Rada, and R. Kosova, “Fuzzy clustering approaches based on numerical optimizations of modified objective functions,” Algorithms, vol. 18, no. 6, p. 327, 2025.
- [16] J. Rada-Vilela, “The FuzzyLite libraries for fuzzy logic control,” 2018. [Online]. Available: <https://fuzzylite.com>

Synthesis of a Fuzzy Rule Base Based on the Results of Fuzzy Clustering of Regular Expressions Using Differential Evolution Algorithms

N.A. Moroshkin

Abstract – The problem of synthesizing a fuzzy rule base for determining the membership of vector representations of abstract syntax trees of regular expressions to clusters based on the results of fuzzy clustering using the standard fuzzy C-means algorithm and its modifications is considered. The main objects of study are fuzzy rule bases based on the Mamdani and Sugeno algorithms, the parameters of which are optimized using evolutionary algorithms, including the classical Differential Evolution (DE) algorithm and its modifications L-SRTDE and L-SHADE-RSP. The aim of the study is to evaluate the effectiveness of differential evolution algorithms for tuning the parameters of a fuzzy rule base while taking into account the number of clusters and the structural features of regular expressions. A comparative analysis of the results of fuzzy rule base synthesis for different optimization algorithms is carried out. The quality of the synthesized models is evaluated using the F1-score metric. The results of the experimental study confirm the feasibility of using the proposed differential evolution algorithms for tuning the parameters of fuzzy rule bases that provide high accuracy in estimating the membership of regular expressions to clusters.

Keywords – Regular expressions, fuzzy clustering, fuzzy rule base, Mamdani algorithm, Sugeno algorithm, evolutionary algorithm, DE, L-SRTDE, L-SHADE-RSP.

REFERENCES

- [1] L. A. Demidova, V. V. Kirakovsky, and A. N. Pylkin, *Decision-Making under Uncertainty*, 2nd ed., revised. Moscow, 2015.
- [2] K. A. Nikitenko and A. V. Zvyagintseva, "Interpretability of neurosemantic models in their application to applied domains," *Problems of Artificial Intelligence*, no. 2(37), pp. 79–90, 2025.
- [3] L. A. Demidova, Yu. S. Sokolova, and E. V. Nikulchev, "Use of fuzzy clustering algorithms ensemble for SVM classifier development," *International Review on Modelling and Simulations*, vol. 8, no. 4, pp. 446–457, 2015.
- [4] L. A. Demidova and A. V. Gorchakov, "Fuzzy information discrimination measures and their application to low dimensional embedding construction in the UMAP algorithm," *Journal of Imaging*, vol. 8, p. 113, 2022.
- [5] A. V. Leonenkov, *Fuzzy Modeling in MATLAB and fuzzyTECH Environment*. St. Petersburg: BHV-Petersburg, 2005, 736 p.
- [6] N. S. Dybulina and S. A. Tsareva, "A model for evaluating the effectiveness of quality management system processes based on the Mamdani algorithm in the MATLAB environment," *Bulletin of Nosov Magnitogorsk State Technical University*, vol. 23, no. 2, pp. 148–157, 2025.
- [7] M. H. Chan, T. F. Ku, and T. T. Nguyen, "Application of the Sugeno fuzzy inference algorithm for predicting the remaining useful life of aviation radio engineering equipment," *Reliability and Quality of Complex Systems*, no. 2(50), pp. 15–25, 2025.
- [8] A. F. Antipin, E. V. Antipina, and S. A. Mustafina, "A modified fuzzy controller with optimization of technological process operating parameters," *Mechatronics, Automation, Control*, vol. 25, no. 7, pp. 354–361, 2024.
- [9] K. E. Dolzhenko, A. V. Kozhina, D. R. Novikov, and V. R. Kozlov, "Differential evolution algorithm for solving machine learning problems," *Scientific and Technical Bulletin of the Volga Region*, no. 7, pp. 116–118, 2025.
- [10] V. Stanovov and E. Semenkin, "Success rate-based adaptive differential evolution L-SRTDE for CEC 2024 competition," in *Proc. 2024 IEEE Congress on Evolutionary Computation (CEC)*, Yokohama, Japan, Jun. 30–Jul. 5, 2024, pp. 1–8.
- [11] V. Stanovov, S. Akhmedova, and E. Semenkin, "NL-SHADE-RSP algorithm with adaptive archive and selective pressure for CEC 2021 numerical optimization," in *Proc. 2021 IEEE Congress on Evolutionary Computation (CEC)*, Kraków, Poland, Jun. 28–Jul. 1, 2021, pp. 809–816.
- [12] J. Lampinen and I. Zelinka, "Mixed integer-discrete-continuous optimization by differential evolution," in *Proc. 5th Int. Conf. on Soft Computing*, 1999.
- [13] L. A. Demidova and N. A. Moroshkin, "Solving the clustering problem for vector representations of regular expressions," *Bulletin of Voronezh State Technical University*, no. 4(21), pp. 50–59, 2025.
- [14] T. T. Z. Nguyen and L. V. Chernenkaya, "A factor analysis model based on fuzzy C-means clustering," *Proceedings of Tula State University*, no. 1, pp. 329–337, 2023.
- [15] E. Bedalli, S. Hajrulla, R. Rada, and R. Kosova, "Fuzzy clustering approaches based on numerical optimizations of modified objective functions," *Algorithms*, vol. 18, no. 6, p. 327, 2025.
- [16] J. Rada-Vilela, "The FuzzyLite libraries for fuzzy logic control," 2018. [Online]. Available: <https://fuzzylite.com>