

УСОВЕРШЕНСТВОВАНИЕ ИНФОРМАЦИОННО-АНАЛИТИЧЕСКИХ КОМПЛЕКСОВ НА ОСНОВЕ ИЕРАРХИЧЕСКИХ СИСТЕМ НЕЧЕТКОГО ВЫВОДА

Сорокин А. А.¹

(ФГБОУ ВО Астраханский государственный технический университет, Астрахань)

Предложены положения по усовершенствованию вычислительных комплексов, основанных на использовании иерархических систем нечеткого вывода. Положения связаны с предварительной подготовкой данных и последующей интерпретацией результатов. Подготовка данных основана на выполнении операций по нормализации входных параметров к шкале с одинаковым диапазоном значений при условии, что эти значения имеют положительную корреляцию с выходной переменной. Реализация положений основана на использовании кусочных функций. Использование положений позволяет упростить формирование продукционных правил в базе знаний системы нечеткого вывода. Исследование поведения изменения значений выходных оценок в зависимости от количества уровней в иерархической системе нечеткого вывода позволило выявить свойство, связанное с группированием этих оценок в окрестностях численных значений термов, которые описывают выходную переменную. Это позволило отнести иерархическую систему нечеткого вывода к классификатору анализируемых объектов по их состоянию с учетом критериев проводимой оценки. Принадлежность к определенному классу определяется величиной итоговой оценки. Для идентификации групп объектов с близкими свойствами внутри классов предложена совокупность операций, основанная на последовательном использовании алгоритма горной кластеризации, евклидовой метрики и метрики Хаусдорфа. Применение операций позволяет выделять типичных представителей исследуемых классов, а затем определять объекты, которые наиболее близки к ним по параметрам с учетом установленных ограничений на отклонения. Проведенный эксперимент подтвердил работоспособность предложенных положений.

Ключевые слова: иерархическая система нечеткого вывода, нормализация, кластеризация, метрика Хаусдорфа, классификатор объектов, алгоритм горной кластеризации.

1. Введение

В настоящее время одной из тенденций развития систем обработки информации стало решение задач по агрегированию

¹ Александр Александрович Сорокин, к.т.н., доцент (alsorokin.astu@mail.ru).

разнородных данных для идентификации оценки состояния различных объектов. Подобное направление получило название Big Data (аналитика больших данных) [19, 20]. Положения по агрегированию разнородной информации часто используются для реализации систем поддержки принятия решений (СППР), которые получили распространение во время управления хозяйственными процессами различных отраслей экономики. К подобным отраслям относятся банковский сектор, телекоммуникационные, транспортные и другие виды компаний [10, 17, 21].

В [8, 9] описывается, что одной из сложностей идентификации интегральных оценок является реализация оператора агрегирования разнородных параметров. Часто сложности вызваны тем, что величины параметров могут оцениваться по различным метрическим шкалам, иметь разный уровень влияния на итоговый результат. Идентификация одинаковых значений интегральной оценки часто происходит при различных комбинациях значений входных величин. Для разрешения подобных сложностей нашли применение экспертные системы [5, 13]. Методы теории нечетких множеств часто используются как теоретическая основа подобных систем обработки информации. В работах [9, 14] указывается, что одним из ограничений систем нечеткого вывода (СНВ) является усложнение структуры правил и увеличение их числа при возрастании количества входных переменных. В работах [2, 14–16] для упрощения структуры правил предлагается агрегирование групп входных переменных при помощи отдельных СНВ с последующим их объединением в единый вычислительный комплекс – иерархическую СНВ (ИСНВ). Как показывает обзор [3, 9, 12], специфика функционирования ИСНВ требует дальнейшего развития положений по предварительной подготовке исходных данных и постобработки полученных значений итоговой оценки.

Цель работы: предложение положений по предварительной подготовке и постобработки данных для иерархических систем нечеткого вывода.

2. Анализ особенностей функционирования систем нечеткого вывода

В рамках исследований принимается, что в основе рассматриваемых СНВ заложен алгоритм Сугено (АС). При прочих равных условиях преимущества АС над алгоритмом Мамдани (АМ) с учетом анализа [9, 14] заключается в том, что:

- АС использует более простую операцию дефаззификации, что сокращает затраты времени на выполнение расчетов;
- выходная переменная СНВ на основе АС обладает более широким диапазоном выходных значений;

В процессе функционирования СНВ на основе АС можно выделить следующие основные этапы обработки информации:

- Фаззификация – преобразование переменной из четкой численной формы x в нечеткую χ при помощи оператора преобразования μ_τ – (функции принадлежности) ФП:

$$(1) \quad \chi = \mu_\tau(x).$$

Задача ФП – определить степень принадлежности значения переменной x к элементам терм-множества $T_\chi = \{\tau_\kappa^\chi\}$, где κ – номер термина в терм-множестве, которое описывает переменную χ ; в качестве ФП обычно используются треугольные, трапециевидальные, гауссовы, синглтоны и др. виды функций.

– Агрегирование подусловий правил в зависимости от значений входных переменных в нечеткой форме; в рамках проводимых исследований будут рассматриваться правила типа MISO (англ. multiple input, single output), которые для агрегирования используют операцию логического «И» – « \wedge »:

$$(2) \quad \omega \text{ if } (\chi_1 = \tau_{\kappa_1, \chi_1}) \wedge \dots \wedge (\chi_n = \tau_{\kappa_n, \chi_n}) \Rightarrow (\psi = \tau_{\kappa_\psi}), \xi_\omega,$$

где ω – номер правила; $\langle \chi_1, \dots, \chi_n \rangle$ – множество входных переменных, а n – их порядковый номер во множестве входных переменных $\mathbf{X} = \{\chi_n\}$; $\tau_{\kappa_1, \chi_1}, \dots, \tau_{\kappa_n, \chi_n}$ – значения термов из терм-множеств, используемых для описания входных переменных; i_1, \dots, i_n – номера термов в терм-множествах; ψ^{fuzz} – выходная переменная в нечеткой форме; $\tau_{\theta, \psi}$ – значение термина выходной переменной, θ – номер термина в терм-множестве; ξ – коэффи-

циент доверия (КД). Выходное значение правила φ_ω применительно к АС может быть определено следующим образом:

$$(3) \quad \varphi_\omega = \xi_\omega \cdot \text{prod}(\mu_\tau(x_1), \dots, \mu_\tau(x_n)) = \xi_\omega \cdot \prod_{i=1}^n \mu_\tau(x_i).$$

– Получение численного значения выходной переменной выполняются при помощи соотношения вида

$$(4) \quad y_{out} = \left(\sum_{\omega=1}^{\Omega} \varphi_\omega \zeta_{\tau_{\kappa_{\psi}}} \right) / \sum_{\omega=1}^{\Omega} \varphi_\omega,$$

где $\zeta_{\tau_{\kappa_{\psi}}}$ – численное значение термина $\tau_{\kappa_{\psi}}$, входящего в термножество, которое описывает выходную переменную в СНВ, а Ω – общее количество правил вида (2), входящих в базу знаний (БЗ) СНВ.

Основное ограничение применения СНВ связано с сильным ростом количества правил в зависимости от количества переменных и термов, которые их описывают. Как показывает анализ [2, 3, 12, 14–16], для устранения подобного ограничения предлагается использовать ИСНВ. В процессе формирования ИСНВ множества входных переменных $X^1 = \{x_n\}$ разделяются на m_1 непересекающихся подмножеств:

$$(5) \quad X^1 \supset X_1^1, \dots, X_{m_1}^1 : X_1^1 \cap \dots \cap X_{m_1}^1 = \emptyset.$$

После этого для каждого подмножества переменных формируется отдельная СНВ. В результате формируется множество СНВ первого уровня иерархии – $F_1 = \{f_{m_1}^1\}$. Функционирование СНВ первого уровня иерархии формирует множество переменных X_2 . Эти переменные аналогично (5) разделяются на m_2 непересекающиеся подмножества переменных, для которых формируются СНВ второго уровня иерархии F_2 . Подобные операции выполняются пока не идентифицируется множество содержащее набор целевых переменных Y_v . Так как в рамках исследований рассматриваются СНВ на основе правил вида MISO, то множество Y_v содержит одну выходную переменную y_{out} .

Часто при разработке СНВ эксперту необходимо для каждой переменной предложить термножество, где каждому тер-

му соответствует отдельная ФП со «своими» специфичными характеристиками. В результате семантическое значение двух идентичных термов, используемых для описания переменных, имеющих положительную и отрицательную корреляцию с выходной переменной, будут различаться. Реализация идентичности семантических значений термов подобных переменных приведет к тому, что аналогичные обозначения термов будут находиться в различных частях шкал абсолютных (естественных) значений входных переменных. Подобное порождает неопределенность при формировании БЗ СНВ и может привести к ошибкам при формировании правил. Для сокращения подобных ошибок целесообразно, чтобы у идентичных термов были близкие семантические значения, одинаковая хронология следования термов на шкале значений входных переменных, а сами входные переменные имели положительную корреляцию с выходной переменной. Для этого необходимо приведение различных значений входных переменных к единой шкале оценки.

3. Преобразование значений входных переменных к единой форме представления

Задачей предлагаемых положений по преобразованию входных переменных является обеспечение результата, когда семантические значения термов и хронология шкалы оценки параметра имеют положительную корреляцию с семантическими значениями термов и численными значениями шкалы выходной переменной. Использование подобных положений позволит обеспечить формирование выходной поверхности СНВ, показывающей закономерность взаимодействия входных переменных в монотонно неубывающем виде. Реализация этого возможна за счет нормализации значений входных переменных отложенных на разных метрических шкалах. Примеры операторов нормализации значений входных переменных, подаваемых на вход СНВ, в виде линейной функции описаны в работе [9]:

$$(б) \quad x_{norm} = f_{norm}(x_{фст}) : x_{norm} \in [0, 1],$$

где f_{norm} – оператор нормализации, $x_{фст}$ – фактическое значение параметра. Однако операторы в работе [9] не учитывают нели-

нейного характера закономерности между значением входной и выходной переменной. Для разрешения этого предлагается выполнять нормализацию входных параметров при помощи кусочных функций $f_{norm}^{piec}(x_{ест})$. Подобное можно представить в виде

$$(7) \quad x_{norm} = f_{norm}^{piec}(x_{ест}) = \begin{cases} \varphi_{u_1}(x_{ест}) : x_{ест} \in [x_{u_{1,1}}^{ест}, x_{u_{1,2}}^{ест}), \\ \dots \\ \varphi_{u_{max}}(x_{ест}) : x_{ест} \in [x_{u_{max,1}}^{ест}, x_{u_{max,2}}^{ест}]; \end{cases}$$

где φ – оператор преобразования естественного значения переменной на отрезке $[x_{u_{1,1}}^{ест}, x_{u_{1,2}}^{ест})$ шкалы оценки параметра $x_{ест}$.

Количество участков и их граничные точки, вид функций нормализации на этих участках определяются экспертами или с учетом нормативных документов, регулирующих обращение с оцениваемым объектом. В рамках исследований реализацию подобных операторов предлагается выполнять на основе уравнения прямой, проходящих через две точки, которые характеризуют изменение закономерности влияния значения входной переменной на выходную. В зависимости от специфики переменной предлагаются три основных вида операторов нормализации при помощи кусочных функций:

- для переменных, которые имеют положительную корреляцию на рис. 1а (сплошная линия) или отрицательную корреляцию на рис. 1а (пунктирная линия) с выходной оценкой;
- для переменных, которые имеют точку оптимума для значения выходной оценки, пример показан на рис. 1б.

Примером параметра, имеющего «точку оптимума», является температура помещения, где находится технологическое оборудование. Значение «благоприятной температуры» занимает относительно «узкий» участок на шкале оценки значений этого параметра, а выход за допустимые границы его значения ухудшит эксплуатационные свойства оборудования.

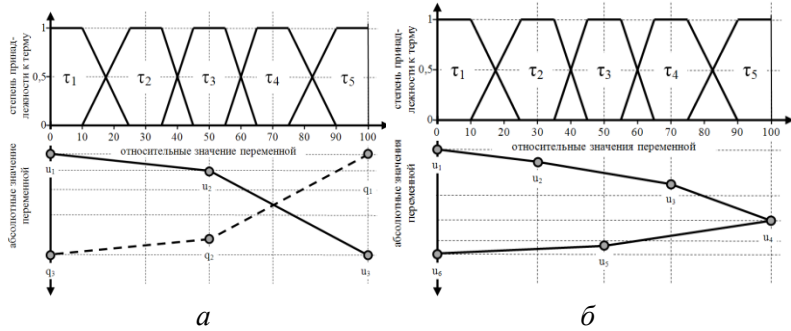


Рис. 1. Вид операторов нормализации значений параметров

Анализ работ [3, 9, 12] показал, что еще одним свойством СНВ является неравномерная чувствительность к изменению значений входных переменных. Результатом подобного становится группировка выходных оценок в окрестностях определенных точек, что приводит к снижению разнообразия оценок и возникновению сложностей идентификации ситуации по значению интегральной оценки. Таким образом, возникает обоснованное предположение, что последовательное применение СНВ еще больше усилит группировку выходных значений в окрестностях определенных точек. Для проверки подобного предположения в следующем разделе описываются результаты эксперимента, посвященного исследованию закономерностей распределения выходных оценок в ИСНВ.

4. Исследование закономерностей функционирования иерархических систем нечеткого вывода

Учитывая результаты сравнения алгоритмов, заложенных в основе функционирования СНВ, дальнейшие рассуждения ведутся из допущения, что при формировании ИСНВ используется алгоритм Сугено. Исследовались ИСНВ следующих видов:

- с однотипными базами правил, в которых семантическое значение выходной переменной равно минимальному семантическому значению входных переменных, далее «СНВ тип 1»;

– с однотипными базами правил, в которых семантическое значение выходной переменной равно максимальному семантическому значению входных переменных, далее «СНВ тип 2»;

Для анализа была построена ИСНВ, состоящая из пяти уровней иерархии. Общее количество входных переменных составило 32, и каждая СНВ, входящая в состав исследуемого вычислительного комплекса, агрегировала две переменные. С учетом соотношения (7) значения входных, промежуточных и выходной переменных задавались в диапазоне от 0 до 100 баллов. Каждый эксперимент включает в себя две серии: ФП для входных переменных СНВ из первой серии показаны на рис. 2а; ФП для входных переменных СНВ из второй серии показаны на рис. 2б.

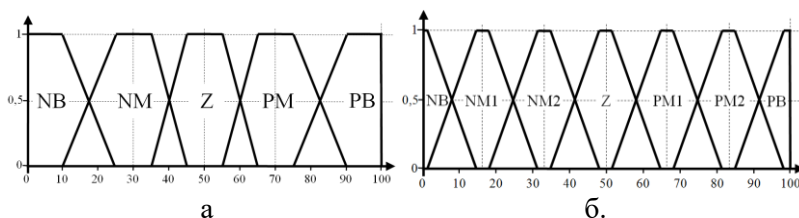


Рис. 2. Функции принадлежности входной переменной

Вид образующих поверхностей элементарных СНВ для:

- первой серии первого эксперимента показан на рис. 3а;
- второй серии первого эксперимента показан на рис. 3б;
- первой серии второго эксперимента показан на рис. 3в;
- второй серии второго эксперимента показан на рис. 3г.

Описание значений переменных проводилось аналогично рекомендациям в работе [6] при помощи следующих термов: NB – низкий; NM – скорее низкий; NM1 – ближе к низкому, чем к среднему; NM2 – ближе к среднему, чем к низкому; Z – средний; PM – скорее высокий; PM1 – ближе к среднему, чем к высокому; PM2 – ближе к высокому, чем к среднему; PB – высокий. Во всех сериях для описания выходных переменных использовались константы, в первой – пять, а во второй – семь. Когда выходная переменная описывается пятью термами, за

каждым термом закреплялись следующие значения: $NB = 0$, $NM = 25$, $Z = 50$, $PM = 75$, $PB = 100$ баллов. Когда выходная переменная описывается семью термами, за каждым термом закреплялись следующие значения: $NB = 0$, $NM1 = 16,67$, $NM2 = 33,33$, $Z = 50$, $PM1 = 66,67$, $PM2 = 83,33$, $PB = 100$ баллов.

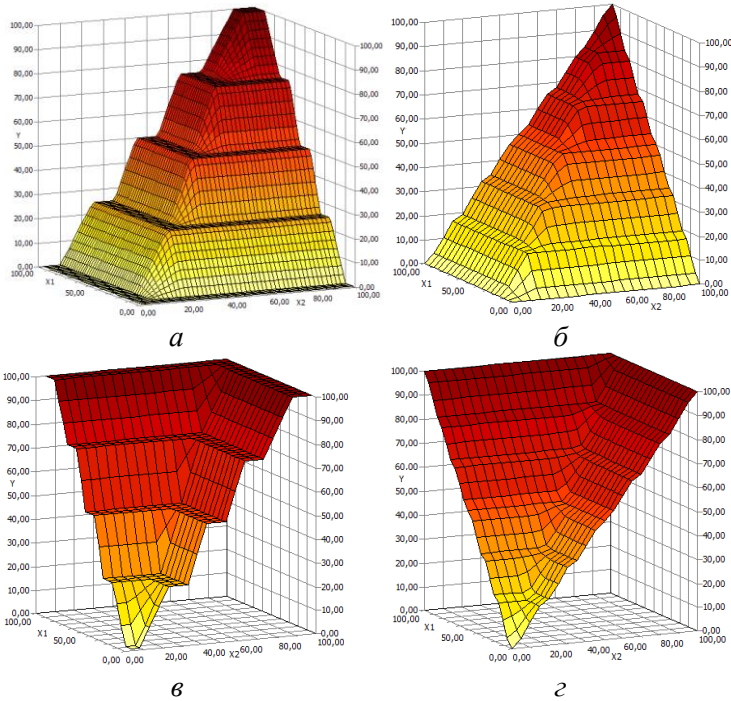


Рис. 3. Поверхности порождаемые СНВ Сугено

В процессе проведения эксперимента на вход первой переменной подавались изменяемые значения в диапазоне от 0 до 100 баллов с шагом 5 баллов, а на другие входы – различные комбинации значений. Комбинации значений включали диапазоны от 0 до 100 баллов с шагом 5 баллов, а также фиксированные значения по 10, 20, 30, 40, 50, 60, 70, 80, 90 и 100 баллов. Учитывая большое количество данных, на рис. 4 приведены за-

кономерности для случаев, когда на входы синхронно подавались значения в диапазоне от 0 до 100 баллов с шагом 5 для:

- первой серии первого эксперимента показан на рис. 4а;
- второй серии первого эксперимента показан на рис. 4б;
- первой серии второго эксперимента показан на рис. 4в;
- второй серии второго эксперимента показан на рис. 4г.

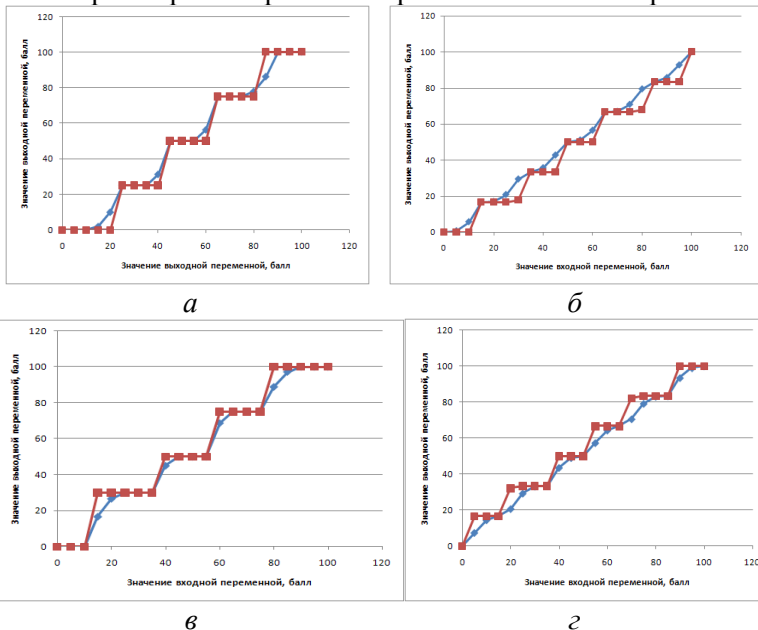


Рис. 4. Закономерности поведения СНВ

Во всех экспериментах заметна тенденция группировки выходных значений в окрестностях значений термов выходной переменной, при увеличении количества иерархий эта тенденция становится более явной. Выходные значения пятого уровня иерархии имеют ступенчатый характер, тогда как выходные значения СНВ первого уровня иерархии имеют более плавный вид. Общий анализ результатов экспериментов позволяет сделать заключение, что:

- при увеличении количества иерархий в вычислительной системе наблюдается группировка выходных значений по определенным областям;
- количество формируемых групп с одинаковыми значениями выходной оценки пропорциональна количеству термов, которые описывают выходную переменную;
- устойчивость результатов группировки наблюдалась после четвертого уровня иерархии;
- области концентрации значений привязаны к значениям термов выходных переменных.

Таким образом, при увеличении уровней иерархии модель, построенная при помощи ИСНВ, теряет чувствительность и ее выходные значения устойчивы относительно значений ФП выходной переменной. В результате ИСНВ фактически становится классификатором оцениваемых объектов по группам состояний. Классификационная способность ИСНВ пропорциональна мощности терм-множества выходной переменной. Общим признаком объектов одного класса является итоговая оценка. Классификатор позволяет разделить исходное множество объектов на подмножества объектов с оценками, соответствующими значениям термов выходной переменной. Подобная классификация позволяет, например, ранжировать объекты по очередности оказания управляющих воздействий, построить график распределения комплексного состояния объектов по определенной территории и решить ряд других задач, в которых важно знать усредненное состояние элементов.

Для оказания управляющего воздействия на определенный элемент необходимо определить, к какой подгруппе внутри выделенного класса этот объект относится. Спецификой подобной задачи является то, что объекты внутри каждой из подгрупп должны незначительно отличаться друг от друга в пределах ограниченных рамок. Следующим этапом исследования стало предложение положений для разделения класса объектов на подгруппы, в которых различие объектов ограничено установленными допусками.

5. Разделение объектов одного класса на подгруппы близких элементов

Как описывается в работах [1, 4, 6, 11] задачи распределения объектов по группам с близкими свойствами при условии неизвестности общего количества групп решаются методами кластеризации. Анализ [1, 4, 6, 11] показал, что алгоритмы кластеризации можно разделить на два вида:

- алгоритмы, в которых проводится расчет «усредненного» центра кластера, с учетом свойств анализируемых объектов, примерами являются алгоритмы k -средних и c -средних;

- алгоритмы, в которых в роли центра кластера выступают определенные вершины анализируемой выборки, а сами объекты выделяются в группу относительно близости к этим вершинам, примером является алгоритм горной кластеризации (АГК) и его модификации.

Учитывая, что в рамках проводимых исследований необходимо выделять объекты близкие к некоторому образцу, который нужно идентифицировать в исходном множестве то применение АГК более целесообразно, так как с его помощью изначально определяется вершина, являющаяся центром некоторой группы.

Работа АГК заключается в выполнении следующей последовательности операций [1, 4, 6, 11]:

1. задается функция $\delta(\lambda_q, \lambda_p)$ с помощью которой рассчитывается расстояние между элементами λ_q и λ_p (описываемые векторами $\langle x_1^q, \dots, x_n^q \rangle$ и $\langle x_1^p, \dots, x_n^p \rangle$), включенными в анализи-

руемое множество $L = \{\lambda_r\} : r = \overline{1, R}$, где R – общее количество элементов множества; часто функция $\delta(\lambda_q, \lambda_p)$ определяется как евклидово расстояние:

$$(8) \quad \delta(\lambda_q, \lambda_p) = \left(\sum_{i=1}^n (x_i^q - x_i^p)^2 \right)^{1/2}.$$

2. Элементы выборки, которые могут быть определены как центры кластеров, обозначаются в виде подмножества

$$N = \{h_t\} : t = \overline{1, T}, \text{ часто справедливо } L = N.$$

3. Для элементов h_t рассчитывается метрика \mathcal{G} , показывающая возможность формирования кластера в ее окрестности:

$$(9) \quad \mathcal{G}_t^1 = \sum_{r=1}^R e^{-\gamma \cdot \delta(h_t, \lambda_r)},$$

где γ – константа, описывающая показатель масштаба расстояния между объектами; часто величина определяется как $1/d_{mean}$, где d_{mean} – среднее расстояние между элементами множества L , в результате формируется множество значений полученных метрик $\{\mathcal{G}_t^1\} = V^1$;

4. Центром первого кластера $h_{c,1}$ назначается элемент h_t с максимальным значением метрики \mathcal{G}_t^1 :

$$(10) \quad h_{c,1} = \arg \max(\{\mathcal{G}_t^1\}).$$

5. При определении центра последующего кластера влияние предыдущего кластера исключается, – это выполняется за счет повторного расчета метрик оставшихся вершин h_t , для этого от текущего значения \mathcal{G}_t^1 оставшихся точек h_t отнимается вклад уже определенного центра кластера $h_{c,1}$:

$$(11) \quad \mathcal{G}_t^2 = \mathcal{G}_t^1 - \mathcal{G}_{t, \max}^1 \cdot e^{-\beta \cdot \delta(h_t, h_{c,1})},$$

где β – положительная константа, характеризующая размер кластера (часто полагают $\beta = \gamma$);

6. Центром второго кластера $h_{c,2}$ назначается элемент h_t с максимальным значением \mathcal{G}_t^2 :

$$(12) \quad h_{c,2} = \arg \max(\{\mathcal{G}_t^2\}).$$

7. Операции 5 и 6 выполняются до наступления условия

$$(13) \quad \mathcal{G}_{t, \max}^v < \mathcal{G}_{t, \text{доп}},$$

где v – номер итерации описываемого алгоритма, $\mathcal{G}_{t, \text{доп}}$ – минимально допустимое значение метрики \mathcal{G} (например, $\mathcal{G}_{t, \text{доп}} \approx 0,1 \mathcal{G}_{t, \max}^1$).

Включение определенной вершины в кластер с центром $h_{t,c}$ определяется на основании степени принадлежности μ_{H_t} , которую можно определить соотношением

$$(14) \mu_{H_t} = 1 - \delta_{h_t, h_{c,1}} / \delta_{h_{c,1}, \max},$$

где $\delta_{h_{c,1}, \max}$ – расстояние между центром кластера и максимально удаленной от него вершиной.

Применение описанных операций ограничивается тем, что часто интерес представляет часть объектов, включенных в кластер H_t , при условии, что эти объекты находятся в окрестности центра кластера на расстоянии не более величины ε .

Таким образом, возникает задача выделения из кластера H_t подмножества элементов H_t^ε , для которых расстояние между вершиной h_t и любым элементом выделенного подмножества H_t^ε не должно превышать величины ε . Решение задачи проводится при помощи использования метрики Хаусдорфа [7]. Сущность применения этой метрики заключается в том, что объект $h_t \subset H_t^\varepsilon$ признается элементом подмножества H_t^ε , если расстояние $\delta_{h_t, h_{c,1}}$ меньше наперед заданного числа ε – радиуса окрестности. С учетом [3, 15] предлагается интерпретация метрики Хаусдорфа в виде решения $d(H_t^\varepsilon)$ о включении $d(H_t^\varepsilon) = 1$ или не включении $d(H_t^\varepsilon) = 0$ вершины h_t в подмножество H_t^ε :

$$(15) d(H_t^\varepsilon) = \begin{cases} 1, & \delta_{h_t, h_{c,1}} \leq \varepsilon, \\ 0, & \delta_{h_t, h_{c,1}} > \varepsilon; \end{cases}$$

если условие выполняется для нескольких центров кластеров, то объект ставится в соответствие с кластером, расстояние до центра которого минимально.

После формирования подгрупп H_t^ε для всех центров кластеров может остаться подмножество элементов H_t^Δ , для которых $d(H_t^\varepsilon) = 0$. Для этого подмножества проводится повторное определение центров кластеров при помощи соотношений (8)–(13) с последующим определением вершин, для которых вы-

полнимо условие (15). Выполнение операций остановится, когда все вершины будут распределены по кластерам или останутся вершины в окрестности, которых не будут включены другие вершины. Такие вершины признаются кластерами, которые содержат один элемент. С учетом (8)–(15) разделение групп, полученных ИСНВ можно описать последовательностью операций:

1. Расчет центров кластеров в каждом из выделенных классов объектов, имеющих определенную интегральную оценку.
2. Выделение объектов, которые наиболее близки к рассчитанным центрам кластеров.
3. Расчет центров кластеров для элементов, оставшихся внутри класса.

После формирования групп элементов с близкими свойствами для каждой из групп возможно формирование общего управляющего воздействия в виде сценария σ_t^ε . Совокупность сценариев σ_t^ε для класса объектов образует множество сценариев воздействий Ψ_y на объекты, получивших определенную оценку y_{out} . Объединение совокупностей сценариев Ψ_y образует множество управляющих воздействий Y на все множество L анализируемых элементов:

$$(16) \forall H_t^\varepsilon \stackrel{\text{эксперт}}{\Rightarrow} \sigma_t^\varepsilon : \bigcup_{t=1}^{t_{max}} \sigma_t^\varepsilon = \Psi_y : \bigcup_{y=1}^{y_{max}} \Psi_y = Y.$$

Таким образом, совокупность теоретических положений (1)–(16) формирует аналитический комплекс для обработки информации, позволяющий из множества исследуемых объектов выделять подмножества элементов с одинаковой интегральной оценкой, а затем из каждого подмножества выделять группы объектов с близкими значениями оцениваемых параметров.

6. Исследование работоспособности предложенных положений

В рамках исследований построена ИСНВ для выделения близких по заданным геометрическим размерам элементов вы-

борки. В качестве выборки использована база данных цветов ирисов, содержащаяся в файле «Iris.data» [18]. Атрибуты каждого из объектов и диапазон и значений следующие:

- длина чашелистика (англ. sepal length): $4,3 \div 7,9$ см;
- ширина чашелистика (англ. sepal width): $2,0 \div 4,4$ см;
- длина лепестка (англ. petal length): $1,0 \div 6,9$ см;
- ширина лепестка (англ. petal width): $0,1 \div 2,5$ см.

В рамках эксперимента необходимо сформировать подмножество элементов, которое будет соответствовать условиям: длина чашелистика примерно 5,4 см и более; ширина чашелистика примерно 3,0 см и более; длина лепестка примерно 4,0 см; ширина лепестка примерно 1,2 см. Для описания входных переменных используются ФП, показанные на рис. 1а, агрегирование переменных проводится при помощи СНВ, поверхность БЗ которой показана на рис. 2а. Операторы для преобразования входных переменных от абсолютных значений к относительным с учетом положений (6), (7) и условий задачи имеют вид:

- для длины чашелистика:

$$(17) \quad p_{norm}^{SL} = \begin{cases} p_{1,norm}^{SL} = 0: p_1^{SL} \in [0; 4,7); \\ p_{2,norm}^{SL} = 250 \cdot p_2^{SL} - 1175: p_2^{SL} \in [4,7; 4,9); \\ p_{3,norm}^{SL} = 100 \cdot p_3^{SL} - 440: p_3^{SL} \in [4,9; 5,4); \\ p_{4,norm}^{SL} = 100: p_4^{SL} \in [5,4; +\infty). \end{cases}$$

- для ширины чашелистика:

$$(18) \quad p_{norm}^{SW} = \begin{cases} p_{1,norm}^{SW} = 100: p_1^{SW} \in [0; 3,0); \\ p_{2,norm}^{SW} = -50 \cdot p_2^{SW} + 250: p_2^{SW} \in [3,0; 3,5); \\ p_{3,norm}^{SW} = -150 \cdot p_3^{SW} + 600: p_3^{SW} \in [3,5; 4,0); \\ p_{4,norm}^{SW} = 0: p_4^{SW} \in [4,0; +\infty). \end{cases}$$

– для длины лепестка:

$$(19) \quad p_{norm}^{PL} = \begin{cases} p_{1,norm}^{PL} = 0 : p_1^{PL} \in [0; 2,0); \\ p_{2,norm}^{PL} = \frac{100}{3} \cdot p_2^{PL} - \frac{200}{3} : p_2^{PL} \in [2,0; 3,5); \\ p_{3,norm}^{PL} = 100 \cdot p_3^{PL} - 300 : p_3^{PL} \in [3,5; 4,0); \\ p_{4,norm}^{PL} = -\frac{125}{3} \cdot p_4^{PL} - \frac{800}{3} : p_4^{PL} \in [4,0; 5,2); \\ p_{5,norm}^{PL} = -250 \cdot p_5^{PL} + 1350 : p_5^{PL} \in [5,2; 5,4); \\ p_{6,norm}^{PL} = 0 : p_6^{PL} \in [5,2; +\infty). \end{cases}$$

– для ширины лепестка:

$$(20) \quad p_{norm}^{PW} = \begin{cases} p_{1,norm}^{PW} = 0 : p_1^{PW} \in [0; 0,6); \\ p_{2,norm}^{PW} = 187,5 \cdot p_2^{PW} - 112,5 : p_2^{PW} \in [0,6; 1,0); \\ p_{3,norm}^{PW} = 125 \cdot p_3^{PW} - 50 : p_3^{PW} \in [1,0; 1,2); \\ p_{4,norm}^{PW} = -62,5 \cdot p_4^{PW} + 175 : p_4^{PW} \in [1,2; 1,6); \\ p_{5,norm}^{PW} = -187,5 \cdot p_5^{PW} + 375 : p_5^{PW} \in [1,6; 2,0); \\ p_{6,norm}^{PW} = 0 : p_6^{PW} \in [2,0; +\infty). \end{cases}$$

Анализ соотношений (17)–(20) показал, что до преобразования длина чашелистика имеет положительную корреляцию с итоговой оценкой, ширина чашелистика – отрицательную, а длина и ширина лепестка имеют точки оптимума соответственно в точках 4,0 и 1,2 см.

На первом уровне иерархии СНВ №1 агрегировала параметры p_{norm}^{SL} и p_{norm}^{SW} , СНВ №2 агрегировала параметры p_{norm}^{PL} и p_{norm}^{PW} . На втором уровне иерархии СНВ №3 агрегировала переменные, полученные при помощи СНВ №1 и СНВ №2. Результат распределения выходных значений в соответствии с нумерацией в базе данных показан на рис. 5а. Группировки выходных оценок в порядке их возрастания – на рис. 5б. Анализ рис. 5 показывает выполнение закономерности о группировки значений выходной переменной в определенных областях.

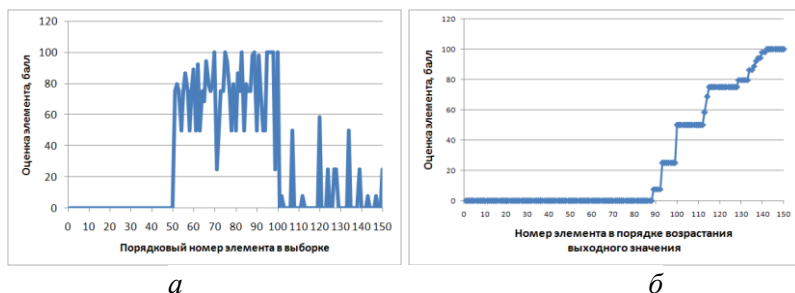


Рис. 5. Выходные значения исследуемой ИСНВ

Перечень объектов выборки, которые наилучшим образом соответствуют задачи классификации, приведены в таблице 1.

Таблица 1. Перечень элементов набравших по 50 баллов выходной оценки

Номер элемента в выборке	Длина чашелисти-ка, см	Ширина чашелистика, см	Длина лепестка, см	Ширина лепестка, см
75	6,4	2,9	4,3	1,3
83	5,8	2,7	3,9	1,2
89	5,6	3	4,1	1,3
95	5,6	2,7	4,2	1,3
96	5,7	3	4,2	1,2
97	5,7	2,9	4,2	1,3
98	6,2	2,9	4,3	1,3
100	5,7	2,8	4,1	1,3
Дисперсия	0,09	0,01	0,02	0,00

Анализ таблицы 1 показывает низкий разброс параметров объектов внутри выборки, что подтверждается малым значением дисперсии. Анализ других выборок показал, что при одинаковом значении интегральной оценки может наблюдаться большой разброс значений параметров. Примером является выборка объектов, набравших 50 баллов, которая приведена в таблице 2. Для выделения групп объектов с близкими параметрами потребовалась их кластеризация.

Таблица 2. Перечень элементов набравших по 50 баллов выходной оценки

Номер элемента в выборке	Длина чашелистика, см	Ширина чашелистика, см	Длина лепестка, см	Ширина лепестка, см
54	5,5	2,3	4	1,3
58	4,9	2,4	3,3	1
61	5	2	3,5	1
63	6	2,2	4	1
72	6,1	2,8	4	1,3
78	6,7	3	5	1,7
80	5,7	2,6	3,5	1
84	6	2,7	5,1	1,6
90	5,5	2,5	4	1,3
93	5,8	2,6	4	1,2
94	5	2,3	3,3	1
107	4,9	2,5	4,5	1,7
134	6,3	2,8	5,1	1,5
Дисперсия	0,33	0,08	0,42	0,08

С использованием положений (8)–(13) определены три вершины, которые можно назначить центрами кластеров, ими стали объекты номер 90, 94, 134. При помощи метрики Евклида было определено расстояние от центров кластеров до остальных объектов класса, результаты расчета представлены в таблице 3.

С учетом (15) принимается, что объект относится к определенному кластеру, если расстояние между ним и центром кластера не более 1 см ($\varepsilon \leq 1$). Если условие выполнимо для нескольких вершин сразу, то объект признается относящимся к тому центру, до которого расстояние минимально. На основании таблицы 2 сформирован перечень групп объектов, наиболее близких к центрам кластеров, который представлен в таблице 4, также в таблице приведены значения дисперсии по каждому из параметров внутри всех полученных групп объектов.

Таблица 3. Расстояния от объектов до центров кластеров

Номер элемента в выборке	Расстояние до вершины №90, см	Расстояние до вершины №94, см	Расстояние до вершины №134, см
54	0,20	0,91	1,46
58	0,97	0,14	2,37
61	0,92	0,36	2,27
63	0,66	1,22	1,38
72	0,67	1,43	1,14
78	1,69	2,60	0,50
80	0,62	0,79	1,79
84	1,26	2,18	0,33
90	0,00	0,93	1,41
93	0,33	1,12	1,26
94	0,93	0,00	2,33
107	0,88	1,41	1,57
134	1,41	2,33	0,00

Таблица 4. Распределение объектов по кластерам

Номер элемента в выборке	Длина чашелистика, см	Ширина чашелистика, см	Длина лепестка, см	Ширина лепестка, см
Кластер №1				
54	5,5	2,3	4	1,3
63	6	2,2	4	1
72	6,1	2,8	4	1,3
80	5,7	2,6	3,5	1
90 (центр кластера)	5,5	2,5	4	1,3
93	5,8	2,6	4	1,2
107	4,9	2,5	4,5	1,7
Дисперсия	0,16	0,04	0,08	0,06
Кластер №2				
58	4,9	2,4	3,3	1
61	5	2	3,5	1
94 (центр кластера)	5	2,3	3,3	1
Дисперсия	0,00	0,04	0,01	0,00

Продолжение таблицы 4

Номер элемента в выборке	Длина чашелистика, см	Ширина чашелистика, см	Длина лепестка, см	Ширина лепестка, см
Кластер №3				
78	6,7	3	5	1,7
84	6	2,7	5,1	1,6
134 (центр кластера)	6,3	2,8	5,1	1,5
Дисперсия	0,12	0,02	0,00	0,01

Анализ таблиц 2 и 4 показывает, что в выделенных группах снизилась дисперсия по всем параметрам, характеризующим состояние объекта. Подобное позволяет прийти к выводу, что внутри групп находятся объекты с более близкими значениями параметров.

7. Заключение

В ходе исследований, предложены положения по предварительной подготовке и постобработке данных для иерархических систем нечеткого вывода. Предварительная подготовка данных заключается в реализации операции нормализации значений входных различных параметров к единой шкале. Реализация положений основана на использовании кусочных функций. После нормализации все параметры приобретают положительную корреляцию с выходной оценкой. Применение предложенных положений упрощает формирование базы правил системы нечеткого вывода, поскольку ее выходная поверхность, интерпретирующая результат взаимодействия входных переменных, принимает монотонно неубывающий характер. Анализ закономерностей изменения оценок в зависимости от количества уровней иерархий в вычислительной системе позволил установить свойство группировки выходных оценок в окрестностях значений термов, описывающих выходную переменную. В результате ИСНВ становится классификатором анализируемых объектов

по группам состояний. Принадлежность к группе определяется величиной итоговой оценки.

Для распределения объектов внутри класса предложена последовательность операций, позволяющих определить объекты, которые могут выступать в роли центра группы, а затем выделять из класса группу объектов, которые наиболее близки к центрам. Реализация положений основана на совместном использовании алгоритма горной кластеризации, Евклидовой метрики и метрики Хаусдорфа. Проведенный эксперимент по оценке элементов множества согласно заранее установленным условиям, подтвердил работоспособность предложенных положений. В результате была сформирована группа объектов, которая полностью соответствует поставленному заданию, эти объекты получили максимальную оценку и были выделены группы объектов, которые частично соответствовали требованиям поставленной задачи. Внутри групп были выделены типичные представители, а потом определены перечни наиболее близких к ним объектов.

Предложенные теоретические положения расширяют возможности использования систем нечеткого вывода в процессе анализа объектов различных видов сложных систем.

Литература

1. АНАНЧЕНКО И.В., ЗУДИЛОВА Т.В., ПОЛИН Я.А., ОСЕТРОВА И.С. *О применимости алгоритмов кластеризации для борьбы со спамом в социальных сетях // Современные наукоемкие технологии.* – 2020. – №4-2. – С. 190–194.
2. ВЕРЕТЁХИН А.В. *Управление уровнем эколого-экономической безопасности промышленного предприятия с применением нечетко-множественного инструментария // Вестник Забайкальского государственного университета.* – 2017. – Т. 23, №6. – С. 128–138. – DOI: 10.21209/2227-9245-2017-23-6-128-138.
3. ЗЁРНЫШКИН А.Е., СИЗОВ А.С., ДОБРИЦА В.П., ГОЛОВИН А.А., ХАЛИН Ю.А. *Модель оценки привлекательности поставщика товарно-материальных ценностей логистиче-*

- ских компаний // Вестник Брянского государственного технического университета. – 2019. – № 3(76). – С. 78–85.
4. ЗУБКОВА Т.М., НАТОЧАЯ Е.Н. *Использование горного алгоритма для оценки затрат при проектировании программных проектов* // Программные продукты и системы. – 2018. – №1. – С. 134–139.
 5. КАРЕЛИН А.Е., БЕРЕЗА А.Н. *Экспертная система для поддержки принятия решений оператора в системе электроснабжения города* // Инженерный вестник Дона. – 2019. – №4(55). – С. 28.
 6. ЛЕОНЕНКОВ А.В. *Нечеткое моделирование в среде MATLAB и fuzzyTech*. – СПб.: БХВ-Петербург, 2005. – 736 с.
 7. ЛЕОНТЬЕВ В.К. *О мерах сходства и расстояниях между объектами* // Журнал вычислительной математики и математической физики. – 2009. – Т. 49, №11. – С. 2041–2058.
 8. НИКИТИН В.В., БОБИН Д.В., НАЗАРОВ А.А. *Моделирование интегрального показателя комплексной оценки объектов исследования* // Вестник Чувашского университета. – 2015. – №3. – С. 172–177.
 9. ПЕГАТ А. *Нечеткое моделирование и управление*. – М.: Бином. Лаборатория знаний, 2017. – 800 с.
 10. ПОЛЕШКИНА И.О., ВАСИЛЬЕВА Н.В. *Технология Blockchain как инструмент управления цепями поставок с участием воздушного транспорта* // Научный вестник Московского государственного технического университета гражданской авиации. – 2020. – Т. 23, №2. – С. 72–86.
 11. ПТАШКО Е.А., УХОБОТОВ В.И. *Автоматическая генерация нечетких правил для управления мобильным роботом с гусеничным шасси на основе числовых данных* // Вестник Южно-Уральского государственного университета. Серия: Вычислительная математика и информатика. – 2017. – Т. 6, №3. – С. 60–72.
 12. САНАЕВА Г.Н., ПРОРОКОВ А.Е., БОГАТИКОВ В.Н., ВЕНТ Д.П. *Иерархическая система нечеткого регулирования процесса получения ацетилена окислительным пиролизом природного газа* // Вестник Астраханского государственного технического университета. Серия: Управление, вычислительная

- техника и информатика. – 2020. – №1. – С. 7–17. – DOI: 10.24143/2072-9502-2020-1-7-17.
13. ЧУВИКОВ Д.А. *Универсальные алгоритмы взаимодействия экспертной системы и системы имитационного моделирования* // Т-Comm: Телекоммуникации и транспорт. – 2017. – Т. 11, №4. – С. 34–40.
 14. ШТОВБА С.Д. *Проектирование нечетких систем средствами MATLAB*. – М.: Горячая линия-Телеком, 2007. – 288 с.
 15. Alrashoud M. *Hierarchical Fuzzy Inference System for Diagnosing Dengue Disease* // IEEE Int. Conf. on Multimedia & Expo Workshops (ICMEW), Shanghai, China, China. – 2019. – P. 31–36. – DOI: 10.1109/ICMEW.2019.00013.
 16. ALRASHOUD M., HAZZA E., ALQAHTANI F., ALHAMMADI M., ABHARI A., GHONEIM A. *Cognitive and Hierarchical Fuzzy Inference System for Generating Next Release Planning in SaaS Applications* // IEEE Access. – 2019. – Vol. 7. – P. 102966–102974. – DOI: 10.1109/ACCESS.2019.2929214.
 17. ELBAHRI F.M., AL-SANJARY O.I., ALI M.A.M., NAIF Z.A., IBRAHIM O., MOHAMMED M. N. *Difference Comparison of SAP, Oracle, and Microsoft Solutions Based on Cloud ERP Systems: A Review* // IEEE 15th Int. Colloquium on Signal Processing & Its Applications (CSPA), Penang, Malaysia, 2019. – P. 65–70. – DOI: 10.1109/CSPA.2019.8695976.
 18. *Iris data set, Machine Learning Repository*. – URL: <https://archive.ics.uci.edu/ml/datasets/iris>
 19. LIU Y., HE J., GUO M., YANG Q., ZHANG X. *An Overview of Big Data Industry in China* // China Communications. – 2014. – Vol. 11(12). – P. 1–10. – DOI: 10.1109/CC.2014.7019834.
 20. LU R., JIN X., ZHANG S., QIU M., WU X. *A Study on Big Knowledge and Its Engineering Issues*. IEEE Transactions on Knowledge and Data Engineering, Vol.: 31, Iss.: 9. – 2019. – P. 1–14. – DOI: 10.1109/TKDE.2018.286686.
 21. PEREIRA S., KARIA D. *AI Use Cases in Operational Support System and Business Support System* // 3rd Int. Conf. on Communication and Electronics Systems (ICCES), Coimbatore, India. – 2018. – P. 15–20. – DOI: 10.1109/CESYS.2018.8724092.

IMPROVEMENT OF INFORMATION-ANALYTICAL COMPLEXES BASED ON HIERARCHICAL SYSTEMS OF FUZZY OUTPUT

Alexandr Sorokin, Astrakhan State Technical University, Astrakhan, Cand.Sc., associate professor (alsorokin.astu@mail.ru).

Abstract: Provisions for the improvement of computing systems based on the use of hierarchical fuzzy inference systems associated with preliminary data preparation and subsequent interpretation of the results are proposed. Data preparation is based on performing operations to normalize the input parameter values to a scale with the same range of values, provided that these values have a positive correlation with the values of the output variable. The implementation of the provisions is based on the use of piecewise functions. The use of provisions makes it possible to simplify the formation of production rules in the knowledge base of the fuzzy inference system. The study of the behavior of changes in estimates depending on the number of levels of hierarchies in the computing system made it possible to identify a property associated with the grouping of output estimates in the vicinity of the values of the terms that describe the output variable. This made it possible to refer the hierarchical system of fuzzy inference to the classifier of the analyzed objects by their state, taking into account the criteria of the assessment. Belonging to a certain class is determined by the value of the integral final grade. For the identify groups of objects with similar properties within classes, a set of operations is proposed based on the sequential use of the mountain clustering algorithm, the Euclidean metric and the Hausdorff metric. The use of operations makes it possible to single out typical representatives of the studied classes, and then determine the objects that are closest to them in terms of parameters, taking into account the established restrictions on deviations. The experiment carried out confirmed the efficiency of the proposed provisions.

Keywords: hierarchical system of fuzzy inference, normalization of input variables, clustering, Hausdorff metric, object classifier, mountain clustering algorithm.

УДК 004.8

ББК 22.18

DOI: 10.25728/ubs.2020.88.5

*Статья представлена к публикации
членом редакционной коллегии Г.А. Угольницким.*

Поступила в редакцию 12.08.2020.

Опубликована 30.11.2020.