

АЛГОРИТМЫ СТРУКТУРНОЙ ИДЕНТИФИКАЦИИ КОМПАКТНЫХ И ТОЧНЫХ НЕЧЕТКИХ СИСТЕМ

Ходашинский Илья Александрович

Д.т.н., профессор, ТУСУР, e-mail: hodashn@rambler.ru

Горбунов Иван Викторович

К.т.н., м.н.с., ТУСУР, e-mail: giv@keva.tusur.ru

Сарин Константин Сергеевич

Ассистент, ТУСУР, e-mail: sks@security.tomsk.ru

Субханкулова София Радиковна

Студентка, ТУСУР, e-mail: sophi1059@gmail.com

Томский государственный университет систем управления и радиоэлектроники
(ТУСУР), 634050 г. Томск, пр. Ленина 40

Аннотация. Представлен алгоритм выбора признаков на основе метода «минный взрыв». Описаны два алгоритма генерации структуры нечетких аппроксиматоров: алгоритм динамического разбиения входного пространства и алгоритм кусочно-линейной инициализации. Рассмотрен алгоритм генерации структуры нечеткого классификатора на основе экстремальных значений таблицы наблюдений. Результаты работы алгоритмов проверены на реальных данных из репозитория KEEL.

Ключевые слова: нечеткие системы типа Такаги-Сугено, структурная идентификация, дискретный алгоритм «минный взрыв», кусочно-линейная инициализация, алгоритм динамического разбиения входного пространства, нечеткий классификатор

Введение. Идентификация системы является важнейшим этапом в процессе нечеткого моделирования. В настоящей работе идентификация рассматривается как этап построения модели, связанный с установлением закономерностей между входными и выходными переменными объекта исследования. Идентификация нечетких систем проводится в два этапа, которые называют структурной и параметрической идентификацией. На этапе структурной идентификации осуществляется определение входных переменных и структурных характеристик нечеткой системы, таких, как число нечетких правил и количество нечетких термов, на которое необходимо разбить входное пространство. На этапе параметрической идентификации в сформированной структуре подбираются параметры системы таким образом, чтобы результаты ее работы был как можно ближе к результатам экспериментальных данных.

Целью статьи является описание алгоритмов выбора информативных признаков (входных переменных) и алгоритмов структурной идентификации компактных и точных нечетких классификаторов и аппроксиматоров.

1. Постановка задачи. Нечеткий аппроксиматор типа Такаги-Сугено задается правилами вида:

ЕСЛИ $x_1=A_{1i}$ AND $x_2=A_{2i}$ AND ... AND $x_n=A_{ni}$ ТО $y = d_{0i} + d_{1i}x_1 + \dots + d_{ni}x_n$,

где n – размерность входного пространства; A_{ji} – лингвистический терм, которым оценивается входная переменная x_j ; выход y задается линейной функцией от входных переменных.

Выход нечеткого аппроксиматора определяет следующее отображение:

$$f(\mathbf{x}; \boldsymbol{\theta}, \mathbf{D}) = \frac{\sum_{i=1}^R \mu_{A_{1i}}(x_1) \cdot \mu_{A_{2i}}(x_2) \cdot \dots \cdot \mu_{A_{ni}}(x_n) \cdot (d_{0i} + d_{1i}x_1 + \dots + d_{ni}x_n)}{\sum_{i=1}^R \mu_{A_{1i}}(x_1) \cdot \mu_{A_{2i}}(x_2) \cdot \dots \cdot \mu_{A_{ni}}(x_n)},$$

где \mathbf{x} – входной вектор, R – число правил; n – количество входных переменных; $\mu_{A_{ji}}$ – функция принадлежности j -ой входной переменной; $\boldsymbol{\theta}$ – вектор параметров функций принадлежности нечеткого аппроксиматора; \mathbf{D} – вектор параметров линейных функций консеквентов правил нечеткого аппроксиматора.

Критерий качества аппроксимации на таблице наблюдений $\mathbf{T}=\{(\mathbf{x}_p; y_p), p = 1, \dots, m\}$ может быть выражен среднеквадратической функцией ошибки:

$$MSE(\boldsymbol{\theta}, \mathbf{D}) = \frac{\sum_{p=1}^m (y_p - f(\mathbf{x}_p; \boldsymbol{\theta}, \mathbf{D}))^2}{m}.$$

Нечеткий классификатор задается правилами следующего вида:

R_{ji} : ЕСЛИ $x_1=A_{j1i}$ И $x_2=A_{j2i}$ И $x_3=A_{j3i}$ И ... И $x_n=A_{jni}$ ТО class= c_j ,

где \mathbf{x} – вектор признаков классифицируемого объекта; c_j – идентификатор j -того класса, $j \in [1, cl]$, A_{jik} – нечеткий терм, характеризующий k -ый признак в ji -ом правиле R_{ji} ($i \in [1, |R_j|]$, $j \in [1, cl]$), R_j – множество правил, относящих наблюдение к классу с идентификатором c_j .

В процессе нечеткой классификации объект относится к каждому классу с определенной степенью принадлежности, вычисленной следующим образом:

$$\beta_j(\mathbf{x}) = \sum_{R_{ji}} \prod_{k=1}^n A_{jik}(x_k), \quad j = 1, 2, \dots, cl.$$

Выходом классификатора является метка класса, определяемая следующим образом:

$$\text{class} = c_{j^*}, \quad j^* = \arg \max_{1 \leq j \leq cl} \beta_j.$$

Нечеткий классификатор может быть представлен функцией $c = f(\mathbf{x}, \boldsymbol{\theta})$, где $\boldsymbol{\theta}$ – вектор, описывающий базу правил.

На множестве обучающих данных (таблице наблюдений) $\{(\mathbf{x}_p; c_p), p = 1, \dots, z\}$ определим единичную функцию

$$\text{delta}(p, \boldsymbol{\theta}) = \begin{cases} 0, & \text{если } c_p = f(\mathbf{x}_p, \boldsymbol{\theta}) \\ 1, & \text{иначе} \end{cases}, \quad p = 1, 2, \dots, z,$$

тогда численный критерий ошибки классификации выражается следующим образом:

$$E(\boldsymbol{\theta}) = \frac{\sum_{p=1}^z \text{delta}(p, \boldsymbol{\theta})}{z}.$$

Целью построения нечётких систем является поиск таких параметров этих систем, которые сводят к минимуму ошибку $E(\theta)$.

2. Выбор признаков. Дискретный алгоритм «Минный взрыв». Алгоритм основан на идее поиска самой взрывоопасной мины, при активации которой всё поле очистилось бы от мин [9]. В данном случае мина будет представлять собой вектор, размерность которого равна числу входных переменных. Значение координаты вектора равно 0, если признак не используется при классификации, иначе – 1.

На начальном этапе определяются параметры алгоритма: \mathbf{X}_0 – начальная точка взрыва (вектор признаков классифицируемого объекта), число осколков N_s , которые разлетаются после взрыва мины и число итераций $maxiter$. Каждая из координат начального вектора равна 1 (считается, что все признаки информативные). Далее вычисляется угол разброса φ , расстояния полёта осколков \mathbf{r} , координаты взорвавшихся мин и координаты новых осколков. Затем определяется, является ли признак информативным. Среди полученных векторов определяется лучшее временное решение с наименьшим значением ошибки классификации. Далее процесс повторяется итерационно, пока количество итераций не достигнет заданного значения.

Псевдокод алгоритма приведен ниже.

Вход: $\mathbf{X}_0, N_s, maxiter$.

Выход: значения: \mathbf{X}_{best} .

$\mathbf{X}[0] := (x_1, x_2, \dots, x_n); iter := 0;$

цикл пока ($iter < maxiter$)

$\mathbf{r}[0] := \text{rand}(\mathbf{max}, \mathbf{min});$

$\varphi := 360 / N_s;$

цикл по i от 1 до N_s

$\mathbf{r}[i] := \mathbf{r}[i - 1] \text{randn};$

$\mathbf{X}_e[i] := \mathbf{r}[i] \cos(\varphi);$

$\mathbf{X}[i] := 1 / (1 + e^{(-\mathbf{X}_e[i])});$

цикл по j от 1 до n

если ($\text{rand}(0,1) < x[i]$) то

$x[i] := 1;$

иначе

$x[i] := 0;$

конец цикла

конец цикла

если ($E(\mathbf{X}[i]) < E(\mathbf{X}_{best})$) то

$\mathbf{X}_{best} := \mathbf{X}[i];$

$iter := iter + 1;$

конец цикла

вывод $\mathbf{X}_{best} := \text{Search_best}(\mathbf{X}[i]).$

3. Динамическое разбиение входного пространства (ДРВП). Алгоритм динамического разбиения входного пространства является алгоритмом структурной идентификации и представляет собой модификацию алгоритма, предложенного в [6]. В данной работе указанный алгоритм адаптирован для систем типа Такаги-Сугено. Идея алгоритма заключается в разбиении входного пространства на нечеткие термы.

На этапе инициализации каждое входное пространство разбивается на один или два нечетких терма таким образом, чтобы ошибка аппроксимации MSE полученной нечеткой системы достигла заданного порога ε . Если достичь порога невозможно, то каждое входное пространство разбивается на два терма. Множество функций принадлежности, на которые разбита переменная i ($i=1, \dots, n$), будем обозначать L_i . Антецеденты находятся как все возможные сочетания функций принадлежности из $\{A_1, A_2, \dots, A_n\}$, обозначим эту процедуру $GetAntecedentParams$. Консеквенты определяются рекуррентным методом наименьших квадратов [7] над теми данными из T , которые подвергаются большему влиянию нечеткого правила, обозначим эту процедуру $GetConsequentParams$.

Далее выполняется итерационный процесс, на каждом шаге которого добавляется новая функция принадлежности в одно из множеств A_1, \dots, A_n и находятся параметры θ и D . Процесс продолжается, пока ошибка аппроксимации MSE нечеткой системы больше заданного порога ε .

Переменная var_worst , в которую добавляется новая функция, находится путем выявления региона reg_worst и переменной в этом регионе, которые вносят больший вклад в ошибку MSE . На пространстве var_worst строится новая функция принадлежности с вершиной a , обозначим эту процедуру $CreateMembership$. Регион входного пространства ограничивается центрами соседних функций принадлежности. На рисунке 1 представлено входное пространство из двух переменных. Каждая переменная разбита на три нечетких терма, что образует четыре региона Reg_1, \dots, Reg_4 . Множество данных таблицы наблюдений, попавших в регион i , будем обозначать M_i .

Псевдокод алгоритма приведен ниже.

Вход: T, ε .

Выход: θ, D .

Инициализация $\{A_1, A_2, \dots, A_n\}$;

$\theta := GetAntecedentParams(\{A_1, A_2, \dots, A_n\})$;

$D := GetConsequentParams(\theta, T)$;

цикл пока ($MSE(\theta, D) > \varepsilon$)

$$reg_worst := \arg \max_{i \in \{1, 2, \dots, CountRegion\}} \frac{\sum_{(x,y) \in M_i} |y - y|}{|M_i|} \cdot \prod_{j=1}^n \frac{L_i^j}{L_j};$$

$$var_worst := \arg \max_{j \in \{1, 2, \dots, n\}} \frac{\sum_{(x,y) \in M_{reg_worst}} |y - y|}{|M_{reg_worst}|} \cdot \frac{L_{reg_worst}^j}{L_j};$$

$$a := \frac{\sum_{(x,y) \in M_{reg_worst}} x_{var_worst} \cdot |y - y|}{\sum_{(x,y) \in M_{reg_worst}} |y - y|};$$

```

 $\mu := \text{CreateMembership}(a, A_{\text{var\_worst}});$ 
 $A_{\text{var\_worst}} := A_{\text{var\_worst}} \cup \{\mu\};$ 
 $\theta := \text{GetAntecedentParams}(\{A_1, A_2, \dots, A_n\});$ 
 $D := \text{GetConsequentParams}(\theta, T);$ 

```

конец цикла

вывод θ, D .

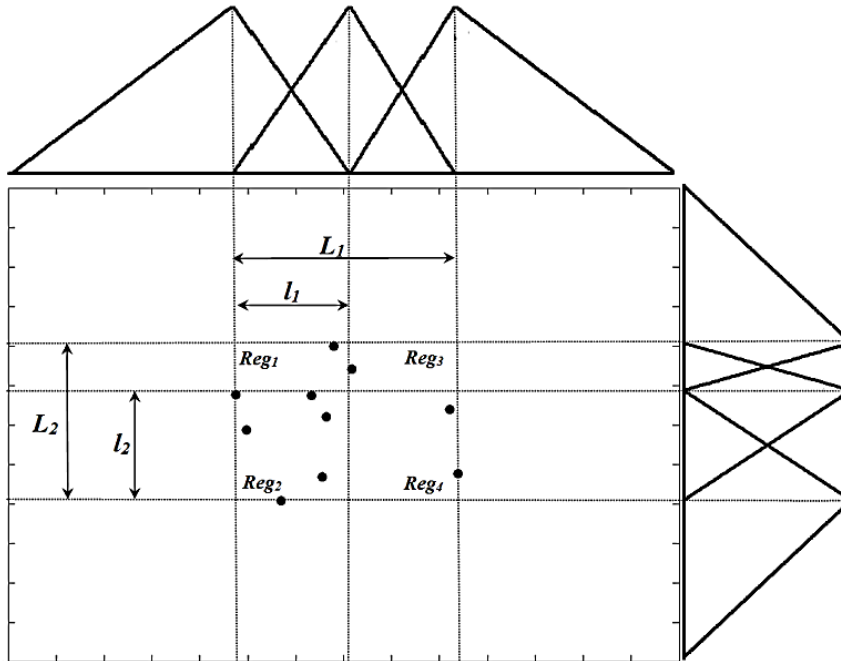


Рис. 1. Построение правил при динамическом разбиении входного пространства

4. Кусочно-линейная инициализация (КЛИ) является алгоритмом структурной идентификации нечетких систем типа Такаги-Сугено. КЛИ относится к алгоритмам кластеризации и отличается от предыдущего тем, что здесь не проводится разбиение входного пространства на нечеткие термы. Вместо этого данные таблицы наблюдений разделяются на группы, называемые кластерами, каждая такая группа будет ассоциироваться с правилом нечеткой системы. Кластеры формируются следующим образом. задается отклонение ϵ . Находится самая удаленная точка (из входных наблюдаемых данных) от начала координат. Относительно этой точки ведется построение кластера: в группу итерационно добавляются ближайшие к этой точке данные до тех пор, пока среднеквадратичная ошибка между выходными данными, входящими в кластер, и линейной регрессией, найденной с помощью данных в кластере рекуррентным методом наименьших квадратов, меньше ϵ . С помощью полученного кластера строится правило.

В этом алгоритме используются функции принадлежности гауссового типа, которые характеризуются двумя параметрами: s – среднее и σ – отклонение. Чтобы найти эти параметры, нужно воспользоваться следующими выражениями:

$$s = \frac{\sum_{k=1}^l x_k}{l}, \quad \sigma = \sqrt{\frac{2}{l} \cdot \sum_{k=1}^l (x_k - s)^2},$$

суммирование здесь ведется по всем данным, входящим в кластер, который представляет правило; l – количество данных в кластере. Консеквент правила соответствует линейной

регрессии, найденной с помощью данных, входящих в кластер, методом наименьших квадратов. Построение кластеров и правил продолжается до тех пор, пока не будут просмотрены все наблюдаемые данные.

Введем следующие обозначения: $Err(C)$ – среднеквадратичная ошибка аппроксимации данных в кластере C гиперплоскостью; $FarPoint(\mathbf{T})$ – самая удаленная точка от начала координат среди множества точек таблицы наблюдений \mathbf{T} ; $FarPoint(P, \mathbf{T})$ – самая удаленная точка от точки P среди точек таблицы наблюдений \mathbf{T} ; $RefreshParams(C, \theta, \mathbf{D})$ – добавление параметров антецедентов и консеквентов нечеткого правила, соответствующего кластеру C , к θ и \mathbf{D} .

Псевдокод алгоритма приведен ниже.

Вход: \mathbf{T}, e .

Выход: θ, \mathbf{D} .

начало цикла ($|\mathbf{T}| > 0$)

$C := \{\emptyset\};$

$P := FarPoint(\mathbf{T});$

$C := C \cup \{P\};$

$\mathbf{T} := \mathbf{T} \setminus \{P\};$

начало цикла ($Err(C) \leq e$ и ($|\mathbf{T}| > 0$))

$p := FarPoint(P, \mathbf{T});$

$C := C \cup \{p\};$

$\mathbf{T} := \mathbf{T} \setminus \{p\};$

конец цикла

$RefreshParams(C, \theta, \mathbf{D});$

конец цикла

вывод θ, \mathbf{D} .

5. Алгоритм генерации структур на основе экстремумов таблицы наблюдений (АГСНК) предназначен для формирования начальной базы правил нечеткого классификатора θ^* , содержащей по одному правилу на каждый класс. Правила алгоритмом формируются на основе экстремальных значений таблицы наблюдений $\{(\mathbf{x}_p; t_p), p = 1, \dots, m\}$ для каждого класса в отдельности [1]. Введем некоторые обозначения: cl – число классов, таблица наблюдений $\{(\mathbf{x}_p; t_p), p = 1, \dots, m\}$, θ^* – база правил классификатора.

Вход: cl .

Выход: База правил классификатора θ^* .

$\theta := \emptyset;$

цикл по j от 1 до m :

цикл по k от 1 до n :

поиск $\min class_{jk} := \min_p(x_{pk});$

поиск $\max class_{jk} := \max_p(x_{pk});$

создание термина A_{jk} , накрывающего интервал $[\min class_{jk}, \max class_{jk}]$;

конец цикла

создание правила R_{1j} на основе термов $A_{jk} (k \in [1, m])$, относящего наблюдение к классу с идентификатором c_j ;

$\theta^* := \theta \cup \{R_{1j}\};$

конец цикла

вывод θ^* .

6. Эксперимент. Эксперимент проводился по схеме кросс-валидации на данных из репозитория KEEL [8]. Описание данных представлено в таблице 1.

Таблица 1. Описание наборов данных

Набор данных	Условное обозначение	Количество записей	Количество признаков	Количество классов
iris	irs	150	4	3
wine	wn	178	13	3
glass	gl	214	9	7
newthyroid	nth	215	5	3
cleveland	cld	297	13	5
monk-2	mnk	306	3	2
bands	bnd	432	6	2
wisconsin	wsn	699	9	2
pima	pm	768	8	2
sonar	snr	208	60	2
vehicle	vhl	846	18	4
coil2000	col	9822	85	2
thyroid	thr	7200	21	3
twonorm	twn	7400	20	2

6.1. Отбор признаков. В ходе эксперимента осуществлялся отбор информативных признаков дискретным алгоритмом «Минного взрыва» (MBA_D). Полученные наборы признаков были проранжированы, из них был выбран набор признаков, на котором проводилась дальнейшая оптимизация параметров нечёткого классификатора с помощью непрерывного алгоритма «Минный взрыв» (MBA_R) [2].

Для генерации структуры нечёткого классификатора использовался АГСНК, в качестве функций принадлежности использовались гауссоиды.

Для дискретного алгоритма «Минный взрыв» было выбрано количество итераций $maxiter = 50$ и количество осколков $N_s = 15$. Для непрерывного алгоритма «Минный взрыв» $maxiter = 100$ и $N_s = 15$. В таблице 2 приведены усредненные значения процента правильной классификации на 11 наборах данных, классифицированных нечеткими классификаторами, структура которых сформирована алгоритмами MBA_R и $MBA_R + MBA_D$, а также результаты работы алгоритмов, представленных в статье [3].

Результаты экспериментов свидетельствуют о том, что уменьшение количества признаков в некоторых случаях ненамного уменьшает процент правильной классификации, а на большинстве наборов данных увеличивает обучающие способности классификатора, делая его абсолютно лучшим (iris) или лучшим относительно классификатора, настроенного только алгоритмом MBA_R (wine, newthyroid, cleveland, bands, vehicle). Положительное влияние оказывает уменьшение признаков и на прогностические способности классификаторов, на наборах данных wine и newthyroid показаны абсолютно лучшие

результаты, относительное улучшение прогностических способностей наблюдается на большинстве наборов данных: iris, glass, cleveland, sonar, bands, vehicle.

Таблица 2. Эффективность нечетких классификаторов

Алгоритм		Наборы данных										
		irs	wn	gl	nth	cld	mnk	bnd	wsn	pm	snr	vhl
		Количество правил										
		3	3	7	3	5	2	2	2	2	2	4
Ant Miner	Обуч.	97.3	99.7	81.5	99.2	60.3	97.2	67.6	92.6	71.9	74.7	59.5
	Тест	96.0	92.1	53.7	90.8	57.5	97.3	59.2	90.4	66.3	71.3	53.1
CORE	Обуч.	95.5	99.1	54.3	92.7	56.3	87.7	66.7	94.7	72.7	53.4	36.5
	Тест	92.7	94.9	45.7	90.8	53.6	88.3	64.2	92.4	73.1	53.4	36.4
HIDER	Обуч.	97.5	97.2	<i>90.1</i>	96.0	<i>82.0</i>	97.2	<i>87.1</i>	97.3	77.8	98.3	84.2
	Тест	96.7	82.6	64.4	90.3	55.9	97.3	62.2	96.1	73.2	52.9	63.1
SGERD	Обуч.	97.3	91.8	53.8	90.2	46.6	80.6	63.8	93.0	73.7	75.7	51.5
	Тест	96.7	87.1	48.3	88.4	44.2	80.7	62.7	92.7	73.7	73.5	51.2
TARGET	Обуч.	93.5	85.2	45.1	88.1	55.8	98.0	71.1	96.1	73.4	76.9	51.6
	Тест	92.9	82.2	44.1	86.8	53.0	96.8	67.3	95.8	73.0	74.6	49.8
MBA _R	Обуч.	97.8	98.7	70.1	98.7	60.7	92.7	70.1	97.0	<i>79.0</i>	78.5	49.4
	Тест	94.4	94.6	62.7	95.8	55.0	92.3	65.3	95.5	74.9	65.6	45.8
MBA _{R+} MBA _D	Обуч.	97.9	99.5	69.8	99.0	61.2	92.1	73.2	96.6	78.8	76.9	50.6
	Тест	95.3	96.8	63.1	96.7	55.6	91.2	67.0	94.7	74.7	66.2	47.4
Номера усечённых признаков		2	6,8,9	8,9	4	2,6,7,9	1,3,4,6	3,6,12,13,14,15	6	4,7	1-16,18-35,37-60	6,16

6.2. Формирование компактной базы правил нечеткого классификатора выполнено алгоритмом АГСНК. В таблице 3 приведены усредненные значения процента правильной классификации на пяти наборах данных, классифицированных нечеткими классификаторами, структура которых сформирована алгоритмом АГСНК и результаты работы алгоритмов, представленных в статье [4], R означает число правил. Жирным шрифтом выделены лучшие результаты классификации на тестовых выборках каждого набора данных. Курсивом выделены лучшие результаты классификации на обучающих выборках. Процент правильной классификации вычислен как разность $100 \cdot (1 - E(\theta))$.

Анализ результатов, представленных в таблице 3, позволяет сделать следующие выводы:

- 1) предложенный алгоритм АГСНК сопоставим по точности с алгоритмами D-MOFARC и FARC-ND как на обучающих, так и на тестовых данных;
- 2) алгоритм АГСНК позволяет получить более компактные базы правил нечеткого классификатора;
- 3) на наборе данных cleveland в случае применения алгоритмов D-MOFARC, FARC-ND наблюдается переобучение.

Таблица 3. Сопоставление результатов с аналогами

Набор данных	Алгоритм								
	АГСНК			D-MOFARC			FARC-HD		
	R	Обуч.	Тест	R	Обуч.	Тест	R	Обуч.	Тест
col	2	94.0	94.0	89.0	94.0	94.0	2.6	94.0	94.0
cld	5	55.1	54.9	45.6	90.9	52.9	42.1	82.2	58.3
nth	3	96.7	88.2	9.5	99.8	95.5	9.6	99.2	94.4
thr	3	99.6	99.3	5.9	99.3	99.1	4.9	94.3	94.1
twn	2	96.7	96.6	10.2	94.5	93.1	60.4	96.6	95.1

6.3. Формирование компактной базы правил нечеткого аппроксиматора

выполнялось двумя алгоритмами: алгоритмом динамического разбиения входного пространства и алгоритмом кусочно-линейной инициализации. В таблице 4 приведены усредненные значения ошибки аппроксимации на 13 наборах данных, полученные нечеткими аппроксиматорами, сформированными алгоритмами КЛИ и ДРВП, а также результаты работы алгоритмов, представленных в статье [5]. При тестировании оценивались такие параметры как количество нечетких правил – R , среднеквадратичная ошибка на обучающей выборке и среднеквадратичная ошибка на тестовой выборке. Пустые значения в таблице говорят о том, что при работе данного алгоритма происходит переполнение памяти, и нормальное завершение алгоритма невозможно. Лучшие результаты ошибки аппроксимации MSE на тестовой выборке выделены жирным шрифтом.

Результаты тестов показали, что предлагаемые алгоритмами структурной идентификации сопоставимы по точности аппроксимации с аналогами, причем количество полученных нечетких правил оказалось значительно меньше. Так же следует отметить, что алгоритмы отработали на всех тестируемых наборах данных. Полученные нечеткие системы прошли только этап структурной идентификации и не были подвергнуты параметрической.

Заключение. В статье описаны методы и алгоритмы выбора признаков и структурной идентификации, позволяющие строить компактные и точные нечеткие системы. Результаты сопоставления полученных нечетких систем с применением предложенных алгоритмов показывают подобные результаты по точности с аналогами при меньшем количестве правил и с использованием меньшего количества входных признаков. Характеристики описанных алгоритмов указывают на целесообразность их применения для построения систем аппроксимации и классификации данных в задачах, где скорость обработки информации не менее важна, чем точность вывода.

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 16-07-00034а и при финансовой поддержке Министерства образования и науки РФ в рамках базовой части государственного задания ТУСУР на 2016 год (проект № 3657).

Таблица 4. Результаты работы алгоритмов идентификации нечетких аппроксиматоров

	ANFIS-SUB			TSK-IRL			LEL-TSK			METSK-HD ^s			ДРВП			КЛИ		
	R	Обуч.	Тест	R	Обуч.	Тест	R	Обуч.	Тест	R	Обуч.	Тест	R	Обуч.	Тест	R	Обуч.	Тест
pla	114	1.011	1.504	21	1.090	1.146	66	1.032	1.188	19.2	1.057	1.136	35.25	1.103	1.128	14.2	1.149	1.168
ele-2	2	8208	8525	262	17024	19786	44.8	2928	3752	36.9	2270	3192	36	9562	9869	3.2	9908	10085
detail	53.8	0.973	1.484	233.2	1.321	1.419	105.2	1.193	1.760	36.8	1.190	1.402	32	1.476	1.479	2.5	1.774	1.776
fried	57.2	0.085	3.158	3055	0.433	1.419	435	0.322	1.070	66	1.075	1.888	72	1.57	1.73	50.2	0.616	0.885
dee	291	3087	2083	3054	0.545	882.01	57.8	0.662	0.682	50.6	0.030	0.103	64	0.087	0.094	1	0.081	0.085
stp	13.2	0.134	0.307				78.9	0.606	0.849	66.4	0.167	0.387	4	2.433	2.548	16.8	0.414	0.527
wan	6	0.639	0.845				123	0.709	1.632	48	0.701	1.189	1	1.285	1.242	12.8	0.918	0.954
wiz	6.2	0.544	0.701				116	0.699	2.227	29.1	0.729	0.944	4	0.749	0.781	4	0.707	0.739
mor	9.4	0.001	0.003				64.3	0.259	0.472	27.2	0.005	0.013	1	0.007	0.006	4.8	0.004	0.004
tre	10	0.009	0.019				63.7	0.267	0.504	28.1	0.017	0.038	1	0.028	0.03	8	0.013	0.024
ca	3	7.14E+11	6.09E+11							32.9	4.376	4.949	1	45.86	47.07	2	33.818	35.851
pole	3	127.40	131.69							46.3	57.96	61.02	1	463.2	465.01	1	463.2	465.01
rum	4	4.482	4.852							63.3	0.267	0.287	1	3.55	3.612	1	3.55	3.612

СПИСОК ЛИТЕРАТУРЫ

1. Ходашинский И.А., Горбунов И.В. Построения нечетких классификаторов на основе алгоритма пчелиной колонии // Материалы Всероссийской конференции с международным участием “Знания – Онтологии – Теории” (ЗОНТ-2011). Новосибирск: Институт математики им. С. Л. Соболева, 2011. Т.2. с. 117–125.
2. Ходашинский И.А., Субханкулова С.Р. Идентификация параметров нечётких систем на основе алгоритма «Минный взрыв» // Информатика и системы управления. 2015. №2 (44). С. 89–98.
3. Alcalá-Fdez J., Fernandez A., Luengo J., Derrac J., García S., Sánchez L., Herrera F. KEEL Data-Mining Software Tool: Data Set Repository, Integration of Algorithms and Experimental Analysis Framework // Journal of Multiple-Valued Logic and Soft Computing. 2011. Vol. 17. P. 255–287.
4. Fazzolari M., Alcalá R., Herrera F. A multi-objective evolutionary method for learning granularities based on fuzzy discretization to improve the accuracy-complexity trade-off of fuzzy rule-based classification systems: D-MOFARC algorithm // Applied Soft Computing. 2014. Vol. 24. P. 470–481. DOI:10.1016/j.asoc.2014.07.019
5. Gacto M.J., Galende M., Alcalá R., Herrera F. METSK-HD^c: A multiobjective evolutionary algorithm to learn accurate TSK-fuzzy systems in high-dimensional and large-scale regression problems // Information Sciences. 2014. Vol. 276. P. 63–79. DOI: [10.1016/j.ins.2014.02.047](https://doi.org/10.1016/j.ins.2014.02.047)
6. Guillaume S. Designing inference systems from data: an interpretability-oriented review // IEEE Transactions on Fuzzy Systems. 2001. Vol. 9. P. 426–443. DOI: 10.1109/91.928739
7. Kalman R.E. A New Approach to Linear Filtering and Prediction Problems // Journal of Basic Engineering. Series D. 1960. Vol. 82. P. 35-45.
8. KEEL-dataset repository [Интернет-портал]. URL: <http://sci2s.ugr.es/keel/datasets.php> (дата обращения: 02.12.2015)
9. Sadollah A., Bahreininejad A., Eskandar H., Hamdi M. Mine blast algorithm: A new population based algorithm for solving constrained engineering optimization problems // Applied Soft Computing. 2013. Vol. 13. P. 2592–2612. DOI: 10.1016/j.asoc.2012.11.026

UDK 004.8

STRUCTURAL IDENTIFICATION OF COMPACT AND ACCURATE FUZZY SYSTEMS

Ilya A. Hodashinsky

Dr., Professor, e-mail: hodashn@rambler.ru

Ivan V. Gorbunov

PhD, e-mail: giv@keva.tusur.ru

Konstantin S. Sarin

Assistant, e-mail: sks@security.tomsk.ru

Sofiya R. Subkhankulova

Student, e-mail: sophi1059@gmail.com

Tomsk State University of Control Systems and Radioelectronics,
40 Lenina Prospect, Tomsk, Russia 634050, Russia

Annotation. The algorithm of feature selection was proposed based on mine blast optimization. Two algorithms are described for a structure generation of fuzzy approximators. The first of them is dynamic partitioning of the input space and the second is piecewise linear initialization. An algorithm are described for structure generation of a fuzzy classifier based on the extreme values of observation table. The results of the algorithms benchmarked on real data from the repository KEEL.

Keywords: Takagi-Sugeno fuzzy systems, structural identification, mine blast algorithm, piecewise linear initialization, dynamic partitioning of the input space, fuzzy classifier

References

1. Hodashinsky I.A., Gorbunov I.V. Postroeniya nechetkikh klassifikatorov na osnove algoritma pchelinoy kolonii // *Materialy Vserossiyskoy konferentsii s mezhdunarodnym uchastiem "Znaniya – Ontologii – Teorii" (ZONT-2011)*. Novosibirsk: Institut matematiki im. S. L. Soboleva, 2011. T.2. S. 117–125. (in Russian).
2. Hodashinsky I.A., Subhankulova S.R. Identifikatsiya parametrov nechetkikh sistem na osnove algoritma «Minnyy vzryv» // *Informatika i sistemy upravleniya*. 2015. №2 (44). S. 89–98. (in Russian).
3. Alcalá-Fdez J., Fernandez A., Luengo J., Derrac J., García S., Sánchez L., Herrera F. KEEL Data-Mining Software Tool: Data Set Repository, Integration of Algorithms and Experimental Analysis Framework // *Journal of Multiple-Valued Logic and Soft Computing*. 2011. Vol. 17. P. 255–287.
4. Fazzolari M., Alcalá R., Herrera F. A multi-objective evolutionary method for learning granularities based on fuzzy discretization to improve the accuracy-complexity trade-off of fuzzy rule-based classification systems: D-MOFARC algorithm // *Applied Soft Computing*. 2014. Vol. 24. P. 470–481. DOI: 10.1016/j.asoc.2014.07.019
5. Gacto M.J., Galende M., Alcalá R., Herrera F. METSK-HDe: A multiobjective evolutionary algorithm to learn accurate TSK-fuzzy systems in high-dimensional and large-scale regression problems // *Information Sciences*. 2014. Vol. 276. P. 63–79. DOI: 10.1016/j.ins.2014.02.047
6. Guillaume S. Designing inference systems from data: an interpretability-oriented review // *IEEE Transactions on Fuzzy Systems*. 2001. Vol. 9. P. 426–443. DOI: 10.1109/91.928739
7. Kalman R.E. A New Approach to Linear Filtering and Prediction Problems // *Journal of Basic Engineering. Series D*. 1960. Vol. 82. P. 35–45.
8. KEEL-dataset repository. URL: <http://sci2s.ugr.es/keel/datasets.php>
9. Sadollah A., Bahreininejad A., Eskandar H., Hamdi M. Mine blast algorithm: A new population based algorithm for solving constrained engineering optimization problems // *Applied Soft Computing*. 2013. Vol. 13. P. 2592–2612. DOI: 10.1016/j.asoc.2012.11.026