УДК 004.85

И.А. Ходашинский, К.С. Сарин

Методика построения компактных и точных нечетких систем типа Такаги–Сугено

Рассматриваются этапы построения нечетких систем типа Такаги–Сугено: формирование исходных данных, генерация структуры нечеткой системы, оптимизация параметров антецедентов и консеквентов нечетких правил. Компромисс между компактностью и точностью модели достигается использованием трех статистических информационных критериев: Akaike, Bayesian, Hannan-Quinn Information Criterion. Приведены результаты экспериментов по аппроксимации реальных данных.

Ключевые слова: нечеткие системы, динамическое разбиение, кусочно-линейная инициализация, алгоритм «кукушкин поиск», рекуррентный метод наименьших квадратов, статистические информационные критерии.

doi: 10.21293/1818-0442-2016-19-1-50-56

Нечеткое моделирование является одним из перспективных методов оценки отношения «вход-выход» в сложных нелинейных системах. Выполняется такое моделирование, как правило, с помощью нечетких систем типа Такаги–Сугено [1]. Нечеткая система представляет реальный объект или процесс в форме, легко понимаемой и интерпретируемой непрофессиональным математиком. В системе нечеткого моделирования собственно знания или модель изучаемого объекта (процесса) отделены от средств манипулирования этими знаниями, что позволяет исследователям ускорить процесс разработки модели, сосредоточившись на вопросе «что делать» вместо вопроса «как делать» [2].

Построение нечетких систем включает четыре основных этапа: 1) формирование исходных данных – задание типа нечеткой системы и сопутствующих этому типу параметров; 2) формирование структуры (грубая настройка); 3) поиск оптимальных значений параметров консеквентов (ТО-частей правил) и параметров функций принадлежности (ФП) антецедентов (ЕСЛИ-части) на основе заданных критериев качества и метода оптимизации выбранного критерия (тонкая настройка); так как антецеденты и консеквенты в системе взаимосвязаны, этот этап является итеративным; 4) проверка правильности построенной системы [3].

Подавляющее большинство методов формирования структуры основано на методах нечеткого кластерного анализа, среди которых наиболее часто применяемыми являются метод нечетких c -средних, алгоритмы Gustafson-Kessel и Gath-Geva, метод нечеткой c -регрессии (FCRM) [4], алгоритм субтрактивной кластеризации [5], а также комбинации указанных методов [6]. Среди перечисленных выше методов FCRM, формирующий форму кластера в виде гиперплоскости, наиболее предпочтителен для решения задач построения нечетких систем типа Такаги–Сугено. Однако реализация алгоритма FCRM для нечеткого моделирования сталкивается со следующими проблемами: 1) длительное время выполнения; 2) неустойчивость к шумам; 3) высокая чувствительность к данным инициализации [7, 8].

Так как оптимизация параметров антецедентов нечетких правил относится к классу NP-трудных задач, выполняется она с использованием различных метаэвристик и их гибридизации [9]. Для оптимизации параметров консеквентов используются различные модификации метода наименьших квадратов.

Проверить правильность построенной системы и избежать переобучения можно, используя процедуру кроссвалидации [10].

В последнее время нечеткие методы моделирования сконцентрированы на проблемах улучшения интерпретируемости нечетких систем без потери точности [11–13]. Лучшая интерпретируемость достигается на компактных базах правил. Однако невозможно построение системы одновременно с высокой степенью точности и компактности, так как эти свойства являются противоречивыми; на практике одно из них преобладает над другим.

При построении компактных и точных нечетких систем задача выбора чаще всего решается либо как двухкритериальная, либо сведением указанных двух критериев к единому критерию. В первом случае различные компромиссы между критериями представлены в виде Парето-оптимального множества сгенерированных нечетких систем [13–19]. Во втором случае из двух критериев либо конструируются один критерий, учитывающий веса каждого [10, 20, 21], либо используются статистические информационные критерии, информационный характер которых связан с концепцией информации Кульбака–Лейблера и которые алгебраически выражаются через сумму меры ошибки аппроксимации и штрафа за количество параметров системы [22–24]. Лучшей считается система с минимальным значением критерия.

Целью настоящей работы является описание оригинальной методики проектирования нечетких систем типа Такаги–Сугено, содержащих оптимальное количество правил и обладающих приемлемой точностью аппроксимации.

Нечеткая система типа Такаги–Сугено

Система типа Такаги–Сугено задается правилами следующего вида [1]:

ЕСЛИ $x_1=A_{1i}$ И ... И $x_n=A_{ni}$ ТО $y=d_{0i}+d_{1i}x_1+...+d_{ni}x_n$.

Выход системы определяет следующее отображение [1]:

$$f(\mathbf{x}; \mathbf{\theta}, \mathbf{D}) = \frac{\sum_{i=1}^R \prod_{j=1}^n \mu_{Aji}(x_j) \cdot (d_{0i} + d_{1i}x_1 + \dots + d_{ni}x_n)}{\sum_{i=1}^R \prod_{j=1}^n \mu_{Aji}(x_j)},$$

где $\mathbf{x}$ – входной вектор; R – число правил; n – количество входных переменных; μ_{Aji} – ФП j -й входной переменной; $\mathbf{\theta}$ – вектор параметров ФП; $\mathbf{D}$ – вектор параметров линейных функций консеквентов правил $\mathbf{D} = [d_{01}, \dots, d_{0R}, d_{11}, \dots, d_{1R}, \dots, d_{n1}, \dots, d_{nR}]^T$.

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

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

Цель построения системы – подбор структуры и параметров, при которых при минимальном количестве правил ошибка $\text{MSE}(\mathbf{\theta}, \mathbf{D})$ сведена к минимуму.

Ниже приведено описание этапов построения нечеткой системы типа Такаги–Сугено.

Формирование исходных данных

Построение нечеткой системы базируется на данных наблюдений, однако в настоящее время не представляется возможным исключить в этом процессе участие специалиста, который решает следующие задачи: 1) задает таблицу наблюдений; 2) указывает входные переменные; 3) определяет тип функций принадлежности.

Формирование структуры

Для формирования структуры системы предлагается использовать два алгоритма: алгоритм динамического разбиения входного пространства (ДРВП), работающий с функциями принадлежности треугольного типа, и алгоритм кусочно-линейной инициализации (КЛИ), использующий гауссовы функции принадлежности.

Алгоритм ДРВП начинает работу с разбиения входного пространства каждой переменной на один или два нечетких термина таким образом, чтобы ошибка аппроксимации MSE сформированной нечеткой системы не превышала заданного порога ε . Если указанное условие не выполняется, то каждое входное пространство разбивается на два термина. Множество нечетких термов (функций принадлежности), на которые разбита i -я переменная ($i = 1, \dots, n$), обозначим Λ_i . База правил задается сочетаниями термов из $\{\Lambda_1, \Lambda_2, \dots, \Lambda_n\}$. Консеквенты определяются рекуррентным методом наименьших квадратов. Далее выполняется итерационный процесс, на каждом шаге которого добавляется новая ФП в одно из множеств $\Lambda_1, \dots, \Lambda_n$ и находятся параметры $\mathbf{\theta}$ и $\mathbf{D}$. Процесс продолжается, пока ошибка аппроксимации MSE больше заданного порога ε [25].

Кусочно-линейная инициализация относится к алгоритмам кластеризации и отличается от описанного выше тем, что здесь не проводится разбиение

входного пространства на нечеткие термы. Вместо этого данные таблицы наблюдений разделяются на группы-кластеры, каждая такая группа будет ассоциироваться с нечетким правилом. Кластеры формируются следующим образом. Задается отклонение e . Находится самая удаленная точка от начала координат (на входных наблюдаемых данных). 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 e . На основе полученного кластера строится нечеткое правило.

Параметры s – среднее и σ – отклонение гауссовых ФП определяются следующим образом [25]:

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

суммирование ведется по данным, входящим в отдельный кластер; l – количество данных в кластере. Консеквенты правил соответствуют уравнениям линейной регрессии, найденной методом наименьших квадратов на входящих в кластер данных. Построение кластеров и правил продолжается до тех пор, пока не будут просмотрены все наблюдаемые данные [25].

Оптимизация параметров

Для поиска оптимальных параметров антецедентов используется градиентный метод и метаэвристический алгоритм «кукушкин поиск» [26].

Идея градиентного метода основана на том, что последующее решение получается из предыдущего движением в направлении, противоположном направлению градиента целевой функции. При оценивании параметров нечетких систем целевой функцией является среднеквадратичная ошибка, а вектор параметров определен на множестве параметров функций принадлежности. Основное достоинство метода – относительно быстрая сходимость, основной недостаток – способность застревать в локальных оптимумах.

Метаэвристический алгоритм «кукушкин поиск» построен на основе имитации способа размножения кукушки, когда она находит недавно построенные гнезда и подкладывает в них свои яйца (заменяет своими), которые в итоге могут быть выкинуты хозяином гнезда [26]. В основе алгоритма лежат три следующих правила: 1) кукушка откладывает по одному яйцу в случайно выбранное гнездо, которое представляет собой решение; 2) часть лучших решений будет перенесена в следующее поколение; 3) количество гнезд фиксировано и есть вероятность того, что хозяин может обнаружить чужое яйцо; в этом случае хозяин может выбросить яйцо из гнезда или вовсе отказаться от гнезда и построить новое на новом месте.

Пошаговая реализация алгоритма представлена ниже.

Шаг 1. Инициализация исходной популяции.

Задается S – размер популяции $\Theta = (\theta^s, s \in [1, S])$, «начальное положение кукушки» выбирается значением θ после построения структуры или после предыдущих этапов оптимизации параметров, являющееся текущим решением θ^{curr} . Задается p – вероятность, с которой гнездо может быть «покинуто» хозяином, т.е. вероятность удаления векторов из множества Θ . Задается количество итераций N в качестве критерия остановки и m – число «худших» решений. Количество генерируемых векторов l полагаем равным S .

Шаг 2. Случайным образом генерируется l векторов решений θ .

Шаг 3. Генерация нового решения на основе полетов Леви.

Выполняется «случайное перемещение кукушки», которое выражено изменением θ^{curr} по закону Леви $\theta^{\text{curr}} = \theta^{\text{curr}} + \text{Levi}$, где Levi – случайный прыжок полета Леви, вычисляемый по правилу $\text{Levi} = \gamma \cdot u / |v|^{1/\beta}$, где γ – коэффициент прыжка полета Леви; параметр β принимает значения из интервала $[1, 2]$; u, v – нормально распределенные величины $v \sim N(0; \sigma_v^2)$, $u \sim N(0; \sigma_u^2)$,

$$\sigma_v^2 = 1, \quad \sigma_u^2 = \left\{ \frac{\Gamma(1+\beta) \sin(\pi\beta/2)}{\Gamma[(1+\beta)/2] \cdot 2^{(\beta-1)/2}} \right\}^{1/\beta} \quad [26],$$

$\Gamma(x)$ – гамма-функция.

Случайным образом выбирается другое решение θ^s из популяции Θ .

Шаг 4. Оценка качества решения.

Если $\text{MSE}(\theta^{\text{curr}}) > \text{MSE}(\theta^s)$, то $\theta^{\text{curr}} = \theta^s$, иначе $\theta^s = \theta^{\text{curr}}$.

Шаг 5. Удаление плохих решений.

Если не выполнено заданное количество итераций N , то выбирается заранее заданное количество m «худших» решений; для каждого из них генерируется случайное число h в диапазоне $[0, 1]$, и если значение h для решения оказывается больше значения заданной вероятности p , гнездо-решение удаляется; вместо удаленного решения формируется новое; l полагается равным количеству уничтоженных векторов-решений, переход на шаг 2.

Иначе выбирается лучшее решение с минимальным значением $\text{MSE}(\theta)$ и выход из алгоритма.

Градиентный метод рано или поздно застревает в локальном минимуме, т.е. при текущей итерации не может улучшить решение, полученное ранее. Решением этой проблемы может быть применение метаэвристического алгоритма «кукушкин поиск», который в силу своего стохастического характера слабо неуязвим для локальных оптимумов. Целесообразно начинать работу гибридного алгоритма с алгоритма «кукушкин поиск», поскольку заранее неизвестно, насколько удачны начальные параметры нечеткой системы.

Статистические информационные критерии

Компактность нечеткой системы определяется количеством самих нечетких правил и количеством antecedentes и консеквентных параметров в пра-

вилах. В нашей работе компактность определяется как сумма количества правил и количества нечетких термов.

Как правило, ошибка MSE на обучающих данных уменьшается по мере усложнения модели, т.е. увеличения числа правил и нечетких термов. Для соблюдения компромисса между компактностью и точностью модели можно использовать три статистических информационных критерия: критерий AIC (Akaike Information Criterion) и BIC (Bayesian Information Criterion), HQC (Hannan-Quinn Information Criterion) [24]. Алгебраически критерии выражаются через сумму меры ошибки модели и штрафа за число параметров модели. Информационный характер критериев связан с концепцией информации Кульбака–Лейблера. Различия между критериями заключаются в определении штрафа (второе слагаемое). С учетом специфики нечеткого моделирования k -я сгенерированная система будет иметь следующие оценки [22]:

$$\text{AIC}(k) = \ln \text{MSE}(\theta, \mathbf{D}) + \frac{2}{m} (m_a + 1 + cR),$$

$$\text{BIC}(k) = \ln \text{MSE}(\theta, \mathbf{D}) + \frac{\ln m}{m} (m_a + 1 + cR),$$

$$\text{HQC}(k) = \ln \text{MSE}(\theta, \mathbf{D}) + \frac{2 \ln(\ln m)}{m} (m_a + 1 + cR),$$

где m – количество наблюдений; m_a – общее число параметров правил; R – число правил; c – эмпирический коэффициент; учитывающий относительную стоимость каждого правила [24].

Согласно приведенным выше критериям из множества сгенерированных систем выбирается та, у которой значение оценки критерия минимально. Таким образом, достигается компромисс между компактностью и точностью системы.

Пошаговое описание выполнения методики

Входными данными для построения нечеткой системы являются данные, сформированные специалистом, выход – компактная и точная система. Ниже приведено пошаговое описание предлагаемой методики.

НАЧАЛО

Шаг 1. Выполнить формирование исходных данных.

Шаг 2. Задать максимальное количество правил R_m .

Шаг 3. ЕСЛИ функция принадлежности гауссового типа,

ТО сгенерировать множество структур алгоритмом КЛИ, содержащим правил не больше R_m , перейти на шаг 4;

ИНАЧЕ ЕСЛИ функция принадлежности треугольного типа,

ТО сгенерировать множество структур алгоритмом ДРВП, содержащим правил не больше R_m , перейти на шаг 4;

ИНАЧЕ ВЫХОД.

Шаг 4. Оценить каждую полученную структуру статистическими информационными критериями AIC, BIC, HQC. Сформировать множество S , состоящее из структур с минимальным значением лю-

бого критерия среди структур с одинаковым количеством нечетких правил.

Шаг 5. Извлечь структуру из S ; $j:=0$; $MSEtr\ a_0:=\infty$.

Шаг 6. $j:=j+1$;

Шаг 7. Оптимизировать параметры антецедентов θ алгоритмом «кукушкин поиск».

Шаг 8. Оптимизировать параметры антецедентов θ алгоритмом градиентного спуска.

Шаг 9. Оптимизировать параметры консеквентов D рекуррентным алгоритмом наименьших квадратов.

Шаг 10. ЕСЛИ $MSEtr_{j-1} > MSEtr_j$, ТО перейти на шаг 6.

Шаг 11. ЕСЛИ S не пусто, ТО перейти на шаг 5.

Шаг 12. Оценить каждую нечеткую систему статистическими информационными критериями.

Шаг 13. ЕСЛИ система после оптимизации параметров показала результаты лучше по двум или более критериям, ТО выбрать оптимизированные системы и перейти на шаг 14.

Шаг 14. Выбрать нечеткую систему путем ранжирования по сумме критериев.

КОНЕЦ

Здесь $MSEtr$ обозначает среднеквадратичную ошибку вывода нечеткой системы на обучающей части таблицы наблюдений. Следует заметить, что все оценки критериев на шагах 4 и 12 проводились на тестовой части таблицы наблюдений, т.е. на данных, которые не входили в этапы формирования структур (шаг 3) и оптимизации параметров (шаги 7–10).

Эксперимент

Исследование методики проводилось при решении задач аппроксимации по схеме кросс-валидации на трех наборах данных из репозитория KEEL (<http://www.keel.es>): ELE-1 (2; 495), Quake (3; 2178), Friedman (5; 1200), здесь в скобках указаны количество входных переменных и общее количество образцов в таблице наблюдений, соответственно.

На этапе структурной идентификации генерируется множество структур алгоритмом кусочно-линейной инициализации. Параметр алгоритма «отклонение» меняется от среднеквадратичной ошибки аппроксимирующей линии регрессии до значения, которому соответствует количество правил в нечеткой системе, равное шести, с интервалом в 1% от ошибки линии регрессии. Лучшие структуры, отобранные по статистическим информационным критериям, приведены в табл. 1, которая представляет собой множество S на шаге 4 методики. Здесь Data – наименование набора данных, N – количество сгенерированных структур, R – количество нечетких правил, $MSEtr$ – среднеквадратичная ошибка вывода нечеткой системы на обучающей части таблицы наблюдений, $MSEst$ – среднеквадратичная ошибка на тестовой части таблицы наблюдений, AIC, BIC и HQC – значения статистических информационных критериев.

На рис. 1 показаны значения критериев и количества правил для нечетких систем, сгенерированных для набора данных Friedman после выполнения шага 3 методики.

Таблица 1

Результаты этапа генерации структур

Data	N	R	MSEtr	MSEst	AIC	BIC	HQC
ELE-1	26	1	389540	476950	13,105	13,256	13,129
		2	497905	457351	13,089	13,366	13,133
		3	489137	442391	13,081	13,483	13,144
		4	433578	404435	13,016	13,545	13,100
		5	353234	413880	13,065	13,719	13,168
		6	355111	404474	13,067	13,847	13,190
Quake	16	1	0,03508	0,03798	–3,262	–3,202	–3,252
		2	0,03507	0,03785	–3,257	–3,146	–3,239
		3	0,03510	0,03790	–3,248	–3,084	–3,222
		4	0,03504	0,03914	–3,207	–2,992	–3,174
		6	0,03510	0,03952	–3,182	–2,863	–3,132
Friedman	30	1	7,262	7,127	1,989	2,136	2,012
		2	6,677	6,410	1,906	2,187	1,950
		3	6,489	6,440	1,933	2,349	1,999
		4	5,912	5,900	1,869	2,419	1,956
		5	4,927	4,721	1,669	2,353	1,777

На этапе оптимизации параметров производился поиск оптимальных параметров для каждой структуры из табл. 1. Значения статистических информационных критериев и среднеквадратичные ошибки полученных систем после настройки параметров приведены в табл. 2 (результат работы шага 12 методики).

По результатам работы двух этапов (см. табл. 1 и 2) сформирован набор нечетких систем с минимальными значениями критериев, и далее путем ранжирования по сумме критериев была выбрана лучшая нечеткая система. Полученные результаты

приведены в табл. 3, здесь поле Σ соответствует сумме рангов трех информационных критериев.

Из табл. 3 видно, что нечеткая система с четырьмя правилами имеет наименьшее значение суммы рангов, равное для набора данных ELE1. Для набора данных Quake наилучшее значение суммы рангов имеет нечеткая система с одним правилом, а для набора данных Friedman – система с тремя правилами. Таким образом, выбранные нечеткие системы будут рекомендованы для аппроксимации указанных наборов данных.

Рис. 1. Значение статистических информационных критериев нечетких систем после этапа формирования структуры для набора данных Friedman. Значения критериев: AIC – *a*; BIC – *б*; HQC – *в*

Таблица 2

Результаты этапа оптимизации параметров

Data	<i>R</i>	MSE _{tra}	MSE _{st}	AIC	BIC	HQC
ELE-1	1	389535	478138	13,239	13,82	13,324
	2	364092	498897	13,423	14,513	13,582
	3	356150	522588	13,611	15,209	13,844
	4	350950	460548	13,626	15,732	13,934
	5	276579	1635828	15,035	17,65	15,417
	6	264097	788690683	21,355	24,478	21,811
Quake	1	0,03508	0,03798	–3,225	–2,992	–3,188
	2	0,03477	0,03803	–3,141	–2,489	–3,038
	3	0,03474	0,03689	–3,213	–2,77	–3,142
	4	0,03455	0,03874	–3,081	–2,219	–2,945
	6	0,03413	0,04148	–2,930	–1,649	–2,727
Friedman	1	7,262	7,127	2,081	2,603	2,162
	2	4,663	4,467	1,722	2,730	1,879
	3	2,513	2,270	1,153	2,646	1,387
	4	2,444	2,401	1,318	3,297	1,627
	5	1,781	2,134	1,308	3,773	1,694

Таблица 3

Выбор лучшей нечеткой системы по сумме рангов

Data	<i>R</i>	MSE _{tra}	MSE _{st}	AIC		BIC		HQC		Σ
				Значение	Ранг	Значение	Ранг	Значение	Ранг	
ELE-1	1	389540	476950	13,105	6	13,256	1	13,129	2	9
	2	497905	457351	13,089	5	13,366	2	13,133	3	10
	3	489137	442391	13,081	4	13,483	3	13,144	4	11
	4	433578	404435	13,016	1	13,545	4	13,1	1	6
	5	353234	413880	13,065	2	13,719	5	13,168	5	12
	6	355111	404474	13,067	3	13,847	6	13,19	6	15
Quake	1	0,03508	0,03798	–3,225	1	–2,992	1	–3,188	1	3
	2	0,03477	0,03803	–3,141	3	–2,489	3	–3,038	3	9
	3	0,03474	0,03689	–3,213	2	–2,77	2	–3,142	2	6
	4	0,03455	0,03874	–3,081	4	–2,219	4	–2,945	4	12
	6	0,03413	0,04148	–2,93	5	–1,649	5	–2,727	5	15
Friedman	1	7,262	7,127	1,989	5	2,136	1	2,012	5	11
	2	4,663	4,467	1,722	4	2,730	3	1,879	4	11
	3	2,513	2,270	1,153	1	2,646	2	1,387	1	4
	4	2,444	2,401	1,318	3	3,297	4	1,627	2	9
	5	1,781	2,134	1,308	2	3,773	5	1,694	3	10

Результаты проведенных экспериментов были сопоставлены с результатами работы четырех методов построения нечетких систем типа Такаги–Сугено [27] и приведены в табл. 4. Краткое описание указанных методов приведено ниже.

ANFIS-SUB – адаптивная нейро-нечеткая система, использующая субтрактивную кластеризацию.

TSK-IRL – метод, использующий индуктивный алгоритм для структурной идентификации нечеткой модели, параметрическая идентификация выполнялась при помощи эволюционного метода.

LEL-TSK – метод, основанный на комбинации индуктивного алгоритма для структурной идентификации и эволюционной стратегии для параметрической идентификации.

METSK-HD^c – масштабируемый двухэтапный метод нечеткого моделирования; структурная идентификация выполняется при помощи генетического алгоритма, для параметрической идентификации используется комбинация генетического алгоритма и фильтра Калмана [27].

Из табл. 4 видно, что при сопоставимой точности (поле MSE_{test}) количество нечетких правил (поле R) в системах, построенных по предлагаемой методике, существенно меньше, что соответствует цели работы – построение компактных и точных нечетких систем.

Таблица 4

Сравнение результатов работы

Data	ANFIS-SUB		TSK-IRL		LEL-TSK		METSK-HD ^c		Наша методика	
	R	MSE _{test}	R	MSE _{test}	R	MSE _{test}	R	MSE _{test}	R	MSE _{test}
ELE-1	28	430000	19	414800	27	480400	11	404400	4	404435
Quake	40	0,31	102	0,046	127	0,0616	18	0,0362	1	0,03798
Friedman	54	6,316	3055	2,838	435	2,14	66	3,776	3	2,270

Заключение

В работе предложена методика построения нечетких систем типа Такаги–Сугено, включающая этапы формирования исходных данных, идентификацию структуры нечеткой системы, оптимизацию параметров antecedentes и консеквентов нечетких правил и выбор оптимальной системы по точности и сложности.

Работоспособность нечетких аппроксиматоров, построенных по разработанной методике, проверена на трех наборах данных из репозитория KEEL.

Последовательная и независимая схема идентификации нечетких систем себя исчерпала, выбрать какой-либо один эффективный алгоритм решения проблемы оптимизации не представляется возможным ввиду сложности и многоплановости самой проблемы, поэтому методика предполагает совместное использование алгоритмов и эвристик одновременно. Дальнейшее развитие предложенной методики заключается в распараллеливании отдельных этапов и процедур.

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 16-07-00034а.

Литература

1. Takagi T. Fuzzy identification of systems and its application to modeling and control / T. Takagi, M. Sugeno // IEEE Transaction Systems, Man and Cybernetics. – 1985. – Vol. 15, No. 1. – P. 116–132.
2. Tron E. Mathematical modeling of observed natural behavior: a fuzzy logic approach / E. Tron, M. Margaliot // Fuzzy Sets and Systems. – 2004. – Vol. 146. – P. 437–450.
3. Ходашинский И.А. Идентификация нечетких систем: методы и алгоритмы // Проблемы управления. – 2009. – № 4. – С. 15–23.
4. Abonyi J. Cluster Analysis for Data Mining and System Identification / J. Abonyi, B. Feil. – Birkhäuser, 2000. – 319 p.
5. Yager R. Generation of fuzzy rules by mountain method / R. Yager, D. Filev // Journal of Intelligent & Fuzzy Systems. – 1994. – Vol. 2. – P. 209–219.
6. Sadrabadi M.R. Identification of the linear parts of nonlinear systems for fuzzy modeling / M.R. Sadrabadi, M.H.F. Zarandi // Applied Soft Computing. – 2011. – Vol. 11. – P. 807–819.
7. T-S fuzzy model identification based on a novel fuzzy c-regression model clustering algorithm / C. Li, J. Zhou, Q. Li,

X. Xiang, X. An // Engineering Applications of Artificial Intelligence. – 2009. – Vol. 22. – P. 646–653.

8. Soltani M. A novel fuzzy c-regression model algorithm using a new error measure and particle swarm optimization / M. Soltani, A. Chaari, F. Ben Hmida // International Journal of Applied Mathematics and Computer Science. – 2012. – Vol. 22, No. 3. – P. 617–628.

9. Handbook of Metaheuristics / Editors M. Gendreau, J.-Y. Potvin. – Springer, 2010. – 669 p.

10. Ishibuchi H. Repeated double cross-validation for choosing a single solution in evolutionary multi-objective fuzzy classifier design / H. Ishibuchi, Y. Nojima // Knowledge-Based Systems. – 2013. – Vol. 54. – P. 22–31.

11. Nguyen C.H. A discussion on interpretability of linguistic rule based systems and its application to solve regression problems / C.H. Nguyen, V.T. Hoang, V.L. Nguyen // Knowledge-Based Systems. – 2015. – Vol. 88. – P. 107–133.

12. Gorzalczyk M.B. A multi-objective genetic optimization for fast, fuzzy rule-based credit classification with balanced accuracy and interpretability / M.B. Gorzalczyk, Rudzinski F. // Applied Soft Computing. – 2016. – Vol. 40. – P. 206–220.

13. Galende-Hernandez M. Complexity reduction and interpretability improvement for fuzzy rule systems based on simple interpretability measures and indices by bi-objective evolutionary rule selection / M. Galende-Hernandez, G.I. Sainz-Palmero, M.J. Fuente-Aparicio // Soft Computing. – 2012. – Vol. 16. – P. 451–470.

14. Pulkkinen P. A Dynamically Constrained Multiobjective Genetic Fuzzy System for Regression Problems / P. Pulkkinen, H. Koivisto // IEEE Transaction on Fuzzy Systems. – 2010. – Vol. 18, No. 1. – P. 161–177.

15. Alcalá R. A Fast and Scalable Multiobjective Genetic Fuzzy System for Linguistic Fuzzy Modeling in High-Dimensional Regression Problems / R. Alcalá, M.J. Gacto, F. Herrera // IEEE Transaction on Fuzzy Systems. – 2011. – Vol. 19, No 4. – P. 666–681.

16. Antonelli M. An efficient multi-objective evolutionary fuzzy system for regression problems / M. Antonelli, P. Ducange, F. Marcelloni // International Journal of Approximate Reasoning. – 2013. – Vol. 54, No 9. – P. 1434–1451.

17. Горбунов И.В. Методы построения трехкритериальных Парето-оптимальных нечетких классификаторов / И.В. Горбунов, И.А. Ходашинский // Искусственный интеллект и принятие решений. – 2015. – № 2. – С. 75–87.

18. Ходашинский И.А. Алгоритмы поиска компромисса между точностью и сложностью при построении нечетких аппроксиматоров / И.А. Ходашинский, И.В. Горбунов // Автометрия. – 2013. – Т. 49, № 6. – С. 51–61.

19. A Review of the Application of Multiobjective Evolutionary Fuzzy Systems: Current Status and Further Direc-

tions / M. Fazzolari, etc. // IEEE Trans. Fuzzy Systems. – 2013. – Vol. 21, No. 1. – P. 45–65.

20. Alcalá-Fdez J. Genetic learning of accurate and compact fuzzy rule based systems based on the 2-tuples linguistic representation / J. Alcalá-Fdez, J. Otero // International Journal of Approximate Reasoning. – 2007. – Vol. 44. – P. 45–64.

21. A new method for designing neuro-fuzzy systems for nonlinear modelling with interpretability aspects / K. Spalka, K. Lapa, A. Przybyl, M. Zalasinski // Neurocomputing. – 2014. – Vol. 135. – P. 203–217.

22. Yen J. Application of Statistical Information Criteria for Optimal Fuzzy Model Construction / J. Yen, L. Wang // IEEE Transaction on Fuzzy Systems. – 1998. – Vol. 6, No 3. – P. 362–372.

23. Zhou S.-M. Low-level interpretability and high-level interpretability: a unified view of data-driven interpretable fuzzy system modeling / S.-M. Zhou, J.Q. Gan // Fuzzy Sets and Systems. – 2008. – Vol. 159. – P. 3091–3131.

24. Ходашинский И.А. Построение компактных и точных нечетких моделей на основе статистических информационных критериев // Информатика и системы управления. – 2014. – № 1 (39). – С. 99–107.

25. Сарин К.С. Три алгоритма генерации структуры нечеткой системы типа Такаги–Сугено / К.С. Сарин, И.А. Ходашинский // Знания-Онтологии-Теории (ЗОНТ-2015): матер. Всерос. конф. с междунар. участием. Российская академия наук, Сибирское отделение; Институт математики им. С.Л. Соболева, 2015. – С. 124–132.

26. Yang X.-S. Engineering optimisation by cuckoo search / X.-S. Yang, S. Deb // Int. J. Mathematical Modelling and Numerical Optimisation. – 2010. – Vol. 1, No. 4. – P. 330–343.

27. METSK-HD^c: A multiobjective evolutionary algorithm to learn accurate TSK-fuzzy systems in high-dimensional and large-scale regression problems / M.J. Gacto, M. Galende, R. Alcalá, F. Herrera // Information Sciences. – 2014. – Vol. 276. – P. 63–79.

Ходашинский Илья Александрович

Д-р техн. наук, профессор каф. комплексной информационной безопасности электронно-вычислительных систем (КИБЭВС) ТУСУРа
Тел.: +7 (382-2) 90-01-11

Эл. почта: hodashn@rambler.ru

Сарин Константин Сергеевич

Ассистент каф. КИБЭВС ТУСУРа

Тел.: +7 (382-2) 70-15-29

Эл. почта: sks@security.tomsk.ru

Hodashinsky I.A., Sarin K.S.

Technique for designing accurate and compact Takagi–Sugeno fuzzy systems

The paper presents the new technique to design accurate and compact Takagi–Sugeno fuzzy systems. The technique includes the following steps: forming the initial data, the generation of fuzzy system structure, optimization of parameters consequent and antecedents of fuzzy rules. The compromise between compactness and accuracy of the model is achieved by using three statistical information criteria: Akaike, Bayesian, Hannan–Quinn Information Criterion. Results of experiments on the approximation of the actual data are delivered. Simulation experiments show validation of the proposed technique.

Keywords: Takagi–Sugeno fuzzy systems, dynamic partitioning, piecewise linear initialization, cuckoo search algorithm, recursive least squares method, statistical information criteria.