Кукса П.П.

**Московский Государственный Технический Университет
им. Н.Э. Баумана**

**E-mail: kouxa@online.ru
WWW: http://www.geocities.com/pkouxa**

Структурная идентификация и оптимизация лингвистических нечетких моделей

Набор решающих правил нечеткой модели может быть получен на основе экспертной информации, данных наблюдений, аналитического описания соотношений между системными переменными. Как структура, так и параметры нечеткой модели могут быть извлечены из данных наблюдений. Данные наблюдений могут использоваться для повышения точности уже имеющейся нечеткой системы (НС) (т.е. для параметрической идентификации нечеткой модели) и автоматической генерации набора правил в форме (1) (т.е. осуществления структурной идентификации нечеткой модели). Нечеткая модель обладает высокой гибкостью и характеризуется большим количеством параметров, которые могут быть настроены: параметры входных нечетких множеств (НМ), определяющие форму и положение функций принадлежности (ФП); параметры алгоритма нечеткого вывода (набор и математическое определение T-, S-, I-, D- операторов и др.), параметры заключений правил. Структура нечеткой модели определяется набором правил и конфигурацией нечетких множеств (их количеством и расположением). Структурная идентификация нечеткой модели включает генерацию набора правил (генерацию структуры) и оптимизацию структуры (индуцированного набора нечетких правил). Оптимизация структуры включает отбор значимых входов и редуцирование нечеткой базы правил.

Лингвистическая нечеткая модель:

ЕСЛИ (U_1 есть A_{11} И … И U_p есть A_{1p}) ТО Y_1 есть B_{11} и … и Y_s есть B_{1s},

…..

ЕСЛИ (U_1 есть A_{n1} И … И U_p есть A_{np}) ТО Y_1 есть B_{n1} и … и Y_s есть B_{ns}. (1)

Предложено [1-7] большое число разнообразных методов решения проблемы идентификации. Превалирующее большинство предложенных методов демонстрируются на простых моделях, а в качестве критериев оценки метода чаще всего используются числовые показатели точности. Интерпретируемость и прозрачность полагается присущей модели по определению, что верно не всегда. В [1] показывается, что для прозрачности модели требуется выполнение многих условий, накладывающих ограничения на форму ФП, конфигурацию НМ, алгоритмы обучения и т.д. Анализ и рассмотрение различных подходов к решению задачи структурной идентификации с точки зрения не только точностных характеристик, но и интерпретируемости и прозрачности получаемых моделей, составляет основную цель данного исследования.

Структурную идентификацию можно рассматривать как комбинаторно-оптимизационную задачу структурного синтеза. В качестве критерия оптимальности может выступать критерий, оценивающий возможности модели по «описанию» данных наблюдений (например, RMSE). Реализация полного перебора всех возможных вариантов НМ при ограничениях (2) (получаемая модель будет прозрачной) на структуру потребует рассмотрения $\binom{m}{q}^p 2^{(q+3)^p}$ различных наборов правил. Полный перебор всех возможных

структур даже при строгой фиксации возможных положений вершин нечетких множеств невозможен уже при средних значениях m, q, p.

$$\forall x_j \in X_j = [l_j, u_j] \ \sum_{i=1}^{q+2} \mu_{i,j}(x_j) = 1, \ j \in \{1,..,p\}$$

$$l_j < x_{0,j} < x_{1,j} < ... < x_{m+1,j} = u_j \qquad (2)$$

$$\mu_{i,j}(x) = \mu_{x_{k_{i-1}}, x_{k_i}, x_{k_{i+1}}}, x \in X_j, \ k_i \in \{1,...,m\}, \ i \in \{1,...,q+2\}, \ k_0 = 0, \ k_{q+3} = m+1$$

Рассмотрим возможные эвристические методы решения задачи структурной идентификации.

I. Индуцирование набора нечетких правил.

В результате анализа выявлены следующие группы методов:

I.1. Группа 1. Методы решетчатого разбиения.

Набор нечетких правил получается в результате разбиения интервалов определения каждой из переменных на подинтервалы (не обязательно такое разбиение имеет физический смысл), на которых задаются нечеткие множества. Декартовы произведения p нечетких множеств A_{i,t_i} порождают в пространстве D входных переменных области, с каждой из которых ассоциируется свое нечеткое правило . Нечеткая посылка правила $R_{t_1,t_2,...,t_p}$ формируется как объединение по И (конъюнкция) p соответствующих нечетких оценок (U_i есть A_{i,t_i}), $i = 1,..., p, t_i \in \{1,...,n_i\}$. В качестве нечеткого заключения $B_{t_1,t_2,...,t_p}$ может быть выбрано либо одноточечное нечеткое множество, либо обычное НМ. В случае одноэлементного НМ после обучения (параметрической оптимизации) нечеткой системы получается спектр значений $\bar{y}_l$ ($l = \overline{1,n}$ где n - количество правил), на основе которого конструируются треугольные ФП, количество которых может быть сокращено за счет объединения близких значений. Если $B_{t_1,...,t_p}$ - обычное множество, то требуются специальные правила направленной адаптации (сдвига, изменения формы и т.д.) нечетких множеств (как выходных, так и входных) в процессе обучения.

Особенности методов данной группы:

1. Недостатки:

-большое количество правил (для НС с p входами и o выходами $n = \prod_{i=1}^{p} n_i \times o$);

-разрешающая способность по данным может оказаться недостаточно высокой, увеличение количества ФП приведет к росту числа правил и возникновению проблем с назначением НМ лингвистических меток;

-часть правил не активизируется в процессе обучения, т.е. возможна ситуация, когда в базе будут присутствовать правила с неопределенными заключениями. Решением проблемы является инициализация неопределенных правил на основе уже инициализированных S:

пусть $N_i = \{r_{i1}, r_{i2}, ...\}$ множество соседних к *i*-му правил (соседнее правило отличается от i-го одним термом) $|N_i| = 2p$, $N_{S_i} = \{i\}, i \in S \, N_{S_i} = N_i, i \notin S$; серия назначений $g^{(0)}(i) = \begin{cases} b_i, i \in S \\ 0 \end{cases}$, $g^{(n+1)}(i) = \frac{1}{2} \inf_{j \in N_{S_i}} g^{(n)}(j) + \frac{1}{2} \sup_{j \in N_{S_i}} g^{(n)}(j)$, $n = 0,1,...$ приводит к инициализации всех правил.

2. Преимущества: 1. Регулярная структура; 2. Небольшое количество различных НМ (лингвистических термов) в модели $\sum_{i=1}^{p} n_i$; 3. Полнота базы нечетких правил $\{R_i\}_1^n$; 4.

Хорошая интерпретируемость получаемой модели (только в том случае, если $\max\{n_1,...,n_p\}$ не более 10), каждое НМ (терм) участвует в $\frac{\prod_1^p n_i}{\sum_1^p n_i}$ правилах.

Динамические методы решетчатого разбиения позволяют изменять топологию пространства D. Для оценки необходимости изменения топологии (добавлении нового НМ (ФП), растягивании, концентрировании имеющихся ФП) можно использовать критерии максимума ошибки, которые локализуют переменную и область, в которую будет добавлено новое НМ. В зонах с малой ошибки количество ФП может быть уменьшено. Недостатком алгоритмов, использующих критерий максимума ошибки, является склонность к излишнему учету выбросов данных, что приводит к концентрированию НМ в окрестностях выбросов. Методы, выделяющие области максимального несоответствия не учитывают текущей конфигурации системы. Для учета конфигурации можно использовать критерий (3) , работающий на уровне правил:

$$CI(i)=\sum_{k=1}^{N}((y_k-R_i)w_i(k))^2 \Rightarrow SCMF(X_j^i)=\sum_{r\in Rc}CI(r),\ Rc=\{R\mid X_j^i\in ant(R)\} \quad (3)$$

Нечеткое разбиение областей определения базовых переменных x_j может осуществляться либо вручную, либо автоматически. В последнем случае возможно равномерное разбиение на q нечетких подмножеств вещественной оси, использование генетических алгоритмов (например, с ЦФ $f=W_cC_s-W_rR_s\to\max$, где R_s-количество правил, C_s-число пар данных, на которых НС дает хороший результат, W_c,W_r- веса) для поиска оптимального числа ФП для каждой переменной, либо использование нечеткой кластеризации (4) вдоль каждой из осей для назначения точкам вещественной оси степеней принадлежности.

$$J(U^j,X_j)=\sum_{k=1}^{N}\sum_{i=1}^{c}(u_{ik}^j)^m\left\|x_j^k-\overline{v_{ij}^1v_{ij}^2}\right\|,\ \text{где}\ \overline{v_{ij}^1v_{ij}^2}=\{x\mid u_{ik}=1,x\in X_j\} \quad (4)$$

I.2. Группа 2. Методы, использующие нечеткую кластеризацию.

Позволяют *одновременно* найти НМ (определить нечеткое разбиение базовых шкал каждой переменной) и индуцировать набор нечетких правил. Кластеризация может осуществляться:

1. в пространстве D - точки, принадлежащие к одному кластеру в пространстве D, могут давать разный результат в Y, т.е. возможны противоречия в заключениях; необходима процедура разрешения конфликтных ситуаций;

2. в пространстве Y – преимуществом является малая размерность пространства; сформированные в Y кластеры проецируются в D; одному выходному кластеру может соответствовать несколько кластеров во входном пространстве;

3. в пространстве D,Y – из-за компенсации расстояний возможно отнесение точек, далеких друг от друга в D и Y, к одному кластеру в D,Y.

Методы проецирования:

1. Проекция кластера C из $U_1\times...\times U_n$ в $U_{i_1}\times..\times U_{i_k}$:

$$\Pr_q C=\int_{X_1\times X_2\times...\times X_p}(\vee_{x_{(q')}}\mu_C(x_1,..,x_p))/(x_{i_1},x_{i_2},..,x_{i_k}),\ \text{где}\ q=(i_1,...,i_k);\ q'=\neg q;$$

2. При проецировании точкам назначается степень принадлежности, совпадающую со степенью принадлежности к кластеру, например, при проецировании из Y в D: $A_1(x_1^k)=...=A_p(x_p^k)=B(y^k)$;

3. Случайная выборка s из множества точек, содержащихся в кластере. Количество необходимых проекций ($p = \sum_{j=1}^{c}(N_j)$) сокращается;

4. Проецируются те точки, для которых u_{ik} близка к 1. Для оставшихся точек назначается степень принадлежности исходя из близости их проекций к интервалам с единичной степенью принадлежности.

Форма кластеров зависит от применяемой меры близости и в большинстве случаев является сферической или гиперэллипсоидальной (гиперэллипсоид произвольной ориентации или расположенный параллельно осям координат).

Особенности нечеткой кластеризации:

- неуправляемый алгоритм (при кластеризации не оптимизируется никакой показатель качества самой НС);

- проблема определения оптимального числа кластеров (решение: 1. использование перед кластеризацией алгоритма, определяющего количество кластеров (например, алгоритма субтрактивной кластеризации); 2. запуск алгоритма кластеризации с разным числом кластеров 3. использование индексов качества разбиения (*S(c)*, *X-B* и т.д.);

- НМ (ФП) порождаются кластерами;

-форма и взаимное расположение ФП может быть необычным (особенно при произвольной ориентации и взаимном расположение кластеров);

- неизбежные потери информации – декартово произведение НМ не дает сам кластер (потери минимизируются при ориентации кластеров параллельно осям координат);

- индивидуальные для каждого нечеткого правила НМ (одно и тоже лингвистическое значение может быть представлено несколькими НМ), что ведет к потере интерпретируемости и прозрачности модели;

- пространство должно быть метрическим;

- для каждой переменной количество НМ в общем случае равно количеству кластеров;

- получаемые при проецировании дискретные ФП должны быть аппроксимированы выпуклой ФП или параметризованной ФП.

Можно выделить следующие методы построения ФП: статистические (анализ получаемых гистограмм); методы, основанные на кластеризации; частные методы проецирования (за вершину ФП принимается проекция центра кластера, диаметр кластера на уровне С определяет форму ФП).

I.3. Группа 3.

Методы этой группы используют многомерные гиперкубические параметризованные ФП (носителем НМ является гиперкуб $H_j = \{A_h, V_j, W_j, B_j(A_h, V_j, W_j)\}$, где V_j - минимальная, а W_j - максимальная точки; A_h - множество данных, принадлежащих H_j).

Особенности: 1. Управляемый алгоритм (есть оптимизация показателя качества самой системы); 2. Отсутствие потерь информации; 3. Кластер (гиперкуб) ориентирован параллельно координатным осям; 4. Пространство должно быть метрическим; 5. Невысокая вычислительная сложность; 6. Нечеткие разбиения определяются автоматически.

I.4. Группа 4. Кластеризация в решетчатой структуре.

Особенности: 1. неуправляемый алгоритм; 2. НМ (ФП) определяют кластеры (как следствие, хорошая интерпретируемость и прозрачность модели) – проекции центра кластера – вершины соответствующих ФП, степень принадлежности к кластеру – Т-норма степеней принадлежности к НМ; 3. Методы данной группы представляют комбинацию нескольких методов и могут рассматриваться как структурно-ориентированные и как гиперкубические методы.

I.5. Группа 5. Структурно-ориентированные методы.

Особенностью методов данной группы является то, что вместо поиска кластеров осуществляется выбор гиперкубов из решетчатой структуры. Нечеткие правила индуцируются селекцией тех ячеек пространства, которые содержат данные. Свойства: 1.

Хорошая интерпретируемость; 2.Низкая вычислительная сложность; 3. Возможность обработки нечисловых данных; 4. Устойчивость к выбросам данных.

I.6. Группа 6.

Построение покрывающего дерева (LOLIMOT [2], CART [3]).

I.7. Группа 7.

Нейро-нечеткие методы [6].

II. Структурная оптимизация.

Структурная оптимизация направлена на повышение компактности центрального решающего блока НС (сокращение количества правил, упрощение их структуры, минимизация числа используемых лингвистических термов и т.д.).

II.A. Отбор значимых входов.

Отбор может быть глобальным или локальным.

Методы селекции входов:

1. Полный перебор – требуется оценить $2^p - 1$ модель.

2. Последовательные методы.

2.1. Инкрементальные методы.

2.1.1. Метод направленного увеличения числа входов до тех пор, пока либо все входы не будут добавлены, либо дальнейшее увеличения числа входов не ведет к улучшению некоторого критерия, оценивающего качество НС. В [7] предложен следующий критерий (4), в соответствии с которым число входов увеличивается пока численное значения критерия не возрастает.

$$RC = \frac{1}{k_A}\sum_{i=1}^{k_A}(y_i^A - \hat{y}_i^{AB})^2 + \frac{1}{k_B}\sum_{i=1}^{k_B}(y_i^B - \hat{y}_i^{BA})^2 \rightarrow \min, \quad (4)$$

где $y_i^A (y_i^B)$ - наблюдаемый выход; $\hat{y}_i^{AB} (\hat{y}_i^{BA})$ - выход модели на данных группы А и В соответственно, после обучения на данных В и А.

Максимальное число рассматриваемых моделей $\frac{p(p+1)}{2} < 2^p - 1.$

2.1.2. На каждом шаге оцениваются p моделей, содержащих 2^j правил (j -номер шага) и j входных переменных; выбирается лучшая M_{j-k_j} по выбранному критерию модель.

2.2. Последовательные методы, работающие в сторону уменьшения числа переменных.

3. Геометрические методы.

3.1. Локальный критерий для правила *i*: $\frac{\Gamma ij}{\Gamma j}$, где $\Gamma_j = X_j$, $\Gamma_{ij} = \{x \mid A_{ij}(x) = 1, x \in X_j\}$, $i \in \{1,..., p\}$, $j \in \{1,...n\}$; глобальный критерий для переменной x_j: $\pi_j = \prod_{i=1}^{n} \frac{\Gamma_{ij}}{\Gamma_j}$.

Чем больше π_j, тем меньше значимость переменной x_j. Селекция переменных по геометрическому критерию может осуществляться на ранних стадиях проектирования, поскольку требуется знание только тех областей на которых соответствующая ФП равна (близка) к 1. При использовании кластеризации пространства Y соответствующие зоны получаются из матрицы $U = [u_{ik}]$. На геометрических критериях также основаны некоторые методы сжатия базы нечетких правил.

3.2. Нечеткие кривые. Для каждой входной переменной строится нечеткая кривая, размах $c_j^{\max} - c_j^{\min}$ которой больше для более значимых переменных.

$$\mu_{jk}(x_j) = \exp(-(x_{jk} - x/b_j)^2) \quad , \quad b_j = \frac{x_j^{\max} - x_j^{\min}}{k} \quad c_{jl} = \frac{\sum_{k=1}^{N} \mu_{jk}(x_{jl}) y_k}{\sum_{k=1}^{N} \mu_{jk}(x_{jl})} \quad ,$$

$l = 1,..., N$

4. Селекция подмножества известной длины.

Требуется оценить $\binom{p}{k}$ моделей. Для ANFIS такая оценка может выполняться на основе первого шага алгоритма обучения (с использованием МНК), поскольку на всех последующих шагах алгоритма используется более медленный метод обратного распространения ошибки BP, приводящий в основном к незначительным модификациям ФП. Выбирается модель, имеющая наименьшую среднеквадратичную ошибку RMSE.

II.B. Сжатие базы правил.

Сжатие базы нечетких правил возможно за счет:

- объединения нечетких множеств, переменных, или слияния кластеров;
- удаления неактивных правил;
- взвешивания правил и удаления правил с весами ниже порогового;
- удаления НМ, близких к универсальным.

Результаты анализа возможных подходов к решению задач структурной идентификации лингвистических нечетких моделей обобщены в табл. 1.

Таблица 1. Сравнение групп методов.

Гр.	количество правил	количество нечетких множеств	интерпретируемость	априорно заданные правила	полнота базы правил	сложность типового алгоритма
1	$n = R^{\max} = \prod_{i=1}^{p} n_i$	$\sum_{i=1}^{p} n_i$	+++[1]	+	+	-
2	$n \geq c$	$\geq cp$	$\pm$	-	-[2]	Nc^2p[3]
3	$\sum_{i=1}^{c} m_i$	$(\sum_{i=1}^{c} m_i) \cdot p$	+++	+	-	Ncp
4	N[4]	$\text{Max} \sum_{i=1}^{p} n_i$	++	+-	+- (Зависит от имеющегося набора данных)	$O(N^2)$
5	$k \leq Q = \min\{N, \prod_{i=1}^{p} n_i\}$	$\sum_{i=1}^{p} n_i$	+++	+	$\min\{N, \prod_{1}^{p} n_i\}$	$O(N^2)$

[1] Предполагается проектирование базы правил вручную.

[2] Для получения полной базы правил нужен алгоритм классификации (например, метод k-ближайших соседей).

[3] Кластеризация Nc^2p (FCM) + индуцирование НМ (npN при использовании кластеризации).

[4] Зависит от имеющегося набора данных.

7	$n = R^{\max} = \prod_{i=1}^{p} n_i$	$\sum_{i=1}^{p} n_i$	от – до +++	+[5]	+	$O(n^3)$

[5] Требуется ввести в алгоритм обучения семантические ограничения.

Литература

1. A. Riid Transparent Fuzzy Systems: Modeling and Control //THESIS ON INFORMATICS AND SYSTEM ENGINEERING, 2002.
2. F. Hoffmann Structure Identification of TSK-Fuzzy System // Fuzzy Sets and Systems, 2001.
3. R. Jang Structure determination in Fuzzy Modeling //Proc. of IEEE IC on Fuzzy Systems, 1994.
4. R. Babuska Data-Driven Fuzzy Modeling, 2001.
5. R. Babuska Fuzzy Systems, Modeling and Identification, 1998.