

Кукса П.П.

Московский Государственный Технический Университет
им. Н.Э. Баумана

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

МЕТОДЫ ОБУЧЕНИЯ НЕЧЕТКИХ СИСТЕМ

Информационно-логические модели систем, построенные на основе теории нечетких множеств и нечеткой логики, находят широкое применение в приложениях управления, принятия решений, системного моделирования, где алгоритмы принятия решений, эвристические алгоритмы управления, знания относительно функционирования рассматриваемого объекта обобщаются и представляются в виде множества нечетких логических правил (1). Качество набора нечетких правил во многом определяется опытом и знаниями экспертов (разработчиков, инженеров, операторов). Процедура трансляции лингвистически заданных правил в объекты теории нечетких множеств характеризуется произвольностью выбора типов функций принадлежности и присвоения значений параметрам этих функций. В результате вид и взаимное расположение функций принадлежности могут быть не вполне адекватны действительности. В то же время характеристики точности и качества нечеткой модели исследуемой системы в значительной степени зависят от параметров формы и взаимного расположения функций принадлежности. Возможным путем преодоления указанных трудностей является применение методов и алгоритмов обучения для осуществления настройки параметров правил и/или адаптации структуры модели с использованием методов структурной самоорганизации и самоадаптации.

$$R^j : IF X_1 \text{ is } A_{1j} \& \dots \& X_n \text{ is } A_{nj} \\ THEN y_1 = f_j^1, \dots, y_m = f_j^m, \quad (1)$$

где f_j^k является нечетким множеством B_j^k , числовым значением b_j^k , вещественной функцией входных переменных $a_{0j}^k + a_{1j}^k x_1 + \dots + a_{nj}^k x_n$, или переменной-индикатором C_j^k принадлежности к k -му классу.

Под обучением нечеткой системы будем понимать процесс формирования базы правил и настройки параметров модели на основе имеющегося набора данных. Параметрическое обучение заключается в нахождение оптимальных значений конкретного множества параметров модели. Алгоритмически параметрическое обучение означает подстройку внутренних параметров модели (параметров формы и положения функций принадлежности, параметров нечетких операций), направленную на улучшение качества модели в смысле принятого критерия. Задача обучения может рассматриваться как задача аппроксимации функциональной зависимости, частично определенной имеющейся выборкой данных. Способность к *обобщению* принципиально имеется у нечетких систем, благодаря свойству универсальной аппроксимации [7]. Под универсальными аппроксимирующими способностями нечетких систем понимается возможность приближения произвольной

функциональной зависимости с любой заданной точностью. Согласно обобщенной аппроксимационной теореме [7], для любой функции g на компактном множестве U и любой $e > 0$, существует нечеткая система $f : U \rightarrow R$, такая что $\sup_{x \in U} |f(x) - g(x)| \leq e$.

Указанное свойство нечетких систем обуславливает потенциальную возможность построение модели, аппроксимирующей поведение исследуемого объекта.

Целью параметрического обучения является нахождение таких значений параметров, при которых *ошибка обобщения модели* $E = \|T - F\| \rightarrow \min$. Прямое измерение указанной ошибки модели на практике не достижимо, поскольку системная функция T при произвольных значениях аргумента во многих случаях не известна. Однако возможно получение оценки ошибки обобщения (2):

$$E_X = \sum_{X \in X} \|F(X) - Y\|, \quad (2)$$

В процессе обучения используются как ошибки на отдельных наборах E_i , так и на всем информационном множестве E (3).

$$\begin{aligned} E_i &= L(\hat{y}_i - y_{di}) \\ E &= \frac{1}{N} \sum_{i=1}^N E_i \end{aligned} \quad (3)$$

В качестве функции потерь $L(\cdot)$ обычно используется квадратичная функция: $L(\cdot) = \frac{1}{2}(\dots)^2$. Задачей параметрического обучения является нахождение значений варьируемых параметров w , доставляющих минимум целевой функции ошибки. При этом следует учитывать, что достигнутые при обучении малые погрешности не гарантируют от значительных ошибок при использовании построенной модели.

$$w^* = \arg \min_w E(w) \quad (4)$$

$$\hat{y}_k = F(x_{1k}, x_{2k}, \dots, x_{nk}; w) + \varepsilon_k, \quad k = 1 \dots N, \quad (5)$$

где w - вектор подлежащих оцениванию параметров; ε_k - погрешность аппроксимации; N - длина выборки данных.

Существует несколько вариантов организации процесса обучения. Обучение может осуществляться:

- с адаптацией параметров на каждом шаге в соответствии с ошибкой E_n (обучение в режиме онлайн, оперативное обучение);
- с адаптацией параметров в соответствии с интегральной ошибкой E (автономное обучение).

$$\alpha^{new} = \alpha^{old} - \eta \left. \frac{\partial E}{\partial \alpha} \right|_{\alpha = \alpha^{old}} \quad \text{- оффлайн адаптация} \quad (6)$$

$$\alpha^{new} = \alpha^{old} - \eta \left. \frac{\partial E_i}{\partial \alpha} \right|_{\alpha=\alpha^{old}}, \quad i=1..N \text{ - онлайн адаптация}$$

Вычисление интегральной ошибки обычно осуществляется согласно (7).

$$\frac{\partial E(w)}{\partial \alpha} = \frac{1}{N} \sum_{i=1}^N \frac{\partial E_i(w)}{\partial \alpha} \quad (7)$$

Для организации обучения необходимы два условия: наличие определенной структуры и алгоритма адаптации. Правила адаптации могут быть получены исходя из аналитических выражений модельной функции. Для лингвистических нечетких систем модельная функция имеет вид:

$$\hat{y} = F(x_1, x_2, \dots, x_n, w_a, w_c) = \frac{\sum_{l=1}^R \tau_l(\bar{x}, w_a) b_l(w_c)}{\sum_{l=1}^R \tau_l(\bar{x}, w_a)}, \quad (8)$$

где w_a, w_c -наборы параметров функций принадлежности условия и заключения.

Контролируемое обучение предполагает, что для каждого входного вектора существует целевой вектор, представляющий собой требуемый выход. Предъявляется входной вектор, вычисляется выход модели и сравнивается с соответствующим целевым вектором, погрешность выхода используется для изменения параметров в соответствии с алгоритмом, стремящимся минимизировать ошибку. Векторы обучающего множества предъявляются последовательно, вычисляются ошибки и веса подстраиваются для каждого вектора до тех пор, пока ошибка по всему обучающему массиву не достигнет требуемого уровня.



Рисунок 1. Гибридная стратегия обучения

Шаг обучения η может быть постоянным или переменным. Параметрическое обучение нечетких систем может осуществляться применительно к различным группам параметров. На этапе обучения функции принадлежности предполагаются зависящими от ряда параметров (ширина основания, координаты вершины и т.д.). Предлагаемые алгоритмы обучения [1-6] осуществляют адаптацию параметров функций принадлежности условия, заключения или и условия и заключения. При назначении функций принадлежности экспертом их существенные модификации в процессе обучения не желательны и настройка модели может осуществляться модификацией параметров заключений правил.

Большинство алгоритмов обучения нечетких систем основаны на принципе обратного распространения ошибки и являются *локальными*. В случае глобальных методов оптимизации в основном осуществляется случайный поиск с использованием генетических или эволюционных алгоритмов.

Задача обучения может решаться и как оптимизационная задача с ограничениями на значения изменяемых параметров.

Рассмотрение существующих методов и алгоритмов обучения нечетких систем позволило выделить следующие варианты организации обучения:

- 1) использование только градиентного алгоритма;
- 2) использование МНК;
- 3) использование градиентного алгоритма и однократного применения МНК;
- 4) совместное использование МНК и градиентного алгоритма на каждом шаге обучения.

Выбор алгоритма обучения диктуется требованиями вычислительной сложности, точности, сложности программной или аппаратной реализаций, требованиями к скорости сходимости алгоритма. Для стандартных градиентных методов основным недостатком является их невысокая скорость. Малая фиксированная величина шага негативно влияет на скорость сходимости. Поэтому представляется целесообразным использование шага переменной величины.

При реализации алгоритма адаптации должны быть рассмотрены вопросы дифференцируемости, выбора величины шага, выбраны вид функции принадлежности, порядок метода, определены семантические ограничения.

Решение задачи параметрического обучения нечетких систем первоначально сводилось к нахождению частных производных ошибки (целевой функции) по адаптируемому параметру. Позже было показано, что нечеткая система на вычислительном уровне может быть представлена в виде сетевой структуры, подобной структурам нейронных сетей, к которой могут быть применены нейросетевые методы и алгоритмы обучения. Это устраняет необходимость в повторной разработке алгоритмов обучения применительно к нечетким системам и позволяет использовать широко известные и хорошо изученные нейросетевые алгоритмы; кроме того, многие результаты, полученные для нейронных сетей, автоматически переносятся и на нечеткие системы.

Нечеткому правилу j соответствует набор нечетких подмножеств $L^j = \{A_{ij}\}_{i=1...n}$, определенных на соответствующих универсальных множествах X_i .

Параметрическая оптимизация может осуществляться либо в предположении (9) уникальности нечетких подмножеств, либо в предположении связанности наборов L^j .

$$L^j \cap L^i = \emptyset \Leftrightarrow i \neq j \quad (9)$$

Соответственно, можно выделить два подхода к реализации параметрического обучения: первый ориентирован на систему правил (10), в которой каждое правило имеет свой набор нечетких множеств (IMF-FLS); второй – на систему правил (11), в которой лингвистическим значениям, содержащимся в нескольких правилах, соответствует одно и то же нечеткое множество (FLS).

$$R^{(j)}: \text{IF } x_1 \text{ is } A_{1j} \text{ AND } x_2 \text{ is } A_{2j} \text{ AND } \dots \text{ AND } x_n \text{ is } A_{nj} \text{ THEN } y \text{ is } B_j \quad (10)$$

$$R^{(j)}: \text{IF } x_1 \text{ is } A_{1k(1,j)} \text{ AND } x_2 \text{ is } A_{2k(2,j)} \text{ AND } \dots \text{ AND } x_n \text{ is } A_{nk(n,j)} \text{ THEN } y \text{ is } B_{h(j)}, \quad (11)$$

где $k(i, j) : \{1, 2, \dots, n\} \times \{1, 2, \dots, R\} \rightarrow \{1, 2, \dots, n_i\}$; $h(j) : \{1, 2, \dots, R\} \rightarrow \{1, \dots, H\}$; n - число входных переменных; R - количество правил n_i - число подмножеств в разбиении домена X_i ; H - количество лингвистических значений выходной переменной. Вторым вариантом задания системы правил является более общим: вариант IMF-FLS является частным случаем FLS при $k(i, j) = j$, $h(j) = j$, $n_i = R$, $H = R \forall i \in \{1, 2, \dots, n\} \forall j \in \{1, 2, \dots, R\}$.

Единственность представления лингвистических значений переменных системы (11) обеспечивается за счет правил *связанной* адаптации параметров. Стандартные градиентные алгоритмы параметрической оптимизации, направленные на минимизацию ошибки E_p или E , приводит к получению системы со следующими характеристиками: количество лингвистических термов для переменной совпадает с числом правил, количество различных консеквентов также совпадает с количеством правил. Получение интерпретируемых систем правил требует специальной организации алгоритмов обучения (адаптации) и учета в алгоритмах следующих требований:

- корректность параметров ФП (для треугольных ФП, например, должно выполняться неравенство $m < a < b < c < M$);
- сохранение взаимного расположения (исходного порядка) нечетких подмножеств внутри доменов $((a, b, c) > (a, b, c)_{left}$ и $(a, b, c) < (a, b, c)_{right}$);
- сохранение требуемого уровня перекрытия между смежными нечеткими подмножествами.

Выполнение указанных требований обеспечивает целостность нечеткой модели.

Рассмотрены существующие методы и алгоритмы обучения нечетких систем. Показано, что задача обучения может рассматриваться и решаться как задача аппроксимации. Рассмотрены различные варианты организации процесса обучения. Приведены требования к алгоритмам параметрического обучения с точки зрения обеспечения целостности и интерпретируемости нечетких моделей. Выделено два класса алгоритмов (с независимой и связанной адаптацией) и структур получаемых нечетких моделей.

Литература

1. *P. Dadone* Design Optimization of Fuzzy Logic Systems // Doctor of Philosophy in Electrical Engineering, Blacksburg, Virginia, May 2001. – p. 197.
2. *D. Gorse, A. J. Shepherd, J. G. Taylor* The New ERA in Supervised Learning // Neural Networks, Volume 10, Issue 2, March 1997, pp. 343-352.
3. *P. Arabshahi, R. J. Marks, R. Reed* Adaptation of Fuzzy Inferencing: A Survey // Department of Electrical Engineering, University of Washington FT-10 Seattle, WA 98195 USA.
4. *J.G. Marin-Blazquez, A. Chouchoulas, Q. Shen* An Overview of Recent Approaches for Fuzzy Learning // University of Edinburgh, UK, 2002.
5. *S. Mitra, Y. Hayashi*, Neuro-Fuzzy Rule Generation: Survey in Soft Computing Framework, IEEE Transactions On Neural Networks, vol. 11, no. 3, May 2000.
6. *J. M. Leski* ϵ -Insensitive Learning Techniques for Approximate Reasoning Systems // International Journal of Computational Cognition, Vol.1, No1, pp.21–77, March 2003.
7. *D. Tikk, L. T. Koczy, T. D. Gedeon* A survey on the universal approximation and its limits in soft computing techniques, 2003. p. 22.

8. *Sugeno, M., Yasukawa, T.* A Fuzzy Logic Based Approach to Qualitative Modeling // IEEE Transactions on Fuzzy System, vol. 1 no. 1, 1993, pp. 7-31.
9. *M. Hiirsalmi, E. Kotsakis, A. Pesonen, A. Wolski* Discovery of Fuzzy Models from Observation Data // RESEARCH REPORT TTE1-2000-43, December 2000.