ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ И СИСТЕМЫ

УДК 004.94

В. А. Терехов, К. А. Майков, С. М. Жиряков

ПОСТРОЕНИЕ СЕМАНТИЧЕСКИ СОДЕРЖАТЕЛЬНЫХ ПРАВИЛ ЭКСПЕРТНЫХ СИСТЕМ В ЗАДАЧАХ С ПОВЫШЕННОЙ ТОЧНОСТЬЮ РЕШЕНИЯ

Предлагается метод построения логико-лингвистической модели коррекции нечеткого вывода с учетом прецедентов принятия решения. На основе обобщения базисных функций Фабера — Шаудера разработана модификация алгоритма нечеткого вывода Суджено, позволяющая редуцировать ошибки решения в условиях неизменности семантики начальных определений.

Ключевые слова: *логико-лингвистическая модель, алгоритм Суджено, функции Фабера — Шаудера.*

Введение. В настоящее время одним из средств подготовки специалистов, ориентированных на решение сложных технических задач, являются интерактивные тренажеры, включающие в себя экспертную подсистему, содержащую понятия и слабо формализуемые правила (эвристики), применяемые экспертом-инструктором для проверки основных вариантов решения задачи, формируемых обучаемым. Сдерживающим фактором развития интерактивных тренажеров, позволяющих заменить присутствие инструктора в процессе тренировки, является необходимость удовлетворения противоречивых требований к применяемым методам поиска решения слабо формализуемых задач. С одной стороны, требуется обеспечить возможность уменьшения погрешности решения, возникающей, в частных случаях в области входных данных, вследствие слабой формализации правил его поиска. С другой стороны, необходимо не допускать модификации определений понятий и эвристик, формируемых экспертом-инструктором. Это требование обеспечивает компетентное интерактивное вмешательство системы в тренировочный процесс в целях информирующего или корректирующего воздействия.

Модификация алгоритма нечеткого вывода Суджено. Рассмотрим возможность модификации алгоритма нечеткого вывода Суджено [1] с учетом прецедентов частных решений.

Для представления функциональной зависимости вида $f : R^N \to R^M$ в слабо формализуемой задаче без ограничения общности можно полагать, что логико-лингвистическая модель задачи содержит правила получения решений r_j с ядром $\ker r_j =< A_j \to B_j >$, $A_j = \left\{ (X_k^{j<}, T_{J(j,k)}) \mid k = \overline{1,N},\ X_k^{j<} \in R^N \right\}$; $B_j = (Z^{j>}, T_{J(j,k+1)}),\ Z^{j>} \in R^M$, где $X_k^{j<}$ — определяющие лингвистические переменные, $Z^{j>}$ — переменная вывода [2], $T_{J(j,k)}$ — базовое терм-множество.

Этап логического вывода алгоритма Суджено характеризует значение переменной вывода Z как линейную комбинацию определяющих переменных [3]:

$$z(x_1,...,x_N)=k_0+\sum_{i=1}^{N}k_i x_i\ . \quad (1)$$

В этом случае целевая поверхность отклика выводимой переменной аппроксимируется гиперплоскостью, что может приводить на этапе композиции к получению неприемлемой по величине погрешности [4]. Осуществим модификацию алгоритма Суджено, основываясь на возможности аппроксимации функции произвольного числа переменных суммой значений вкладов каждой определяющей переменной независимо друг от друга, что показано в работах Колмогорова о представлении непрерывных функций нескольких переменных. Тогда требуемая поверхность отклика выводимой переменной может быть представлена выражением [5]

$$z(x_1,..,x_N)=\lim_{L\to\infty}\sum_{l=1}^{L}\sum_{n=1}^{N}\delta_{l,n}(x_n)\,, \quad (2)$$

где l — порядок (уровень) приближения, $\delta_{l,n}(x_n)$ — вклад переменной X_n в значение z на l-м уровне приближения.

Для обеспечения сходимости уравнения (2) необходимо использовать аналогию многомерного обобщения базисных функций системы Фабера — Шаудера [6] и осуществить разбиение пространства $X_1\times...\times X_N$ на зоны решения Ω_d^l, так что

$$\begin{gathered}\Omega_d^l=\bigcup_i\Omega_i^{l+1},\ \Omega_i^{l+1}\cap\Omega_j^{l+1}=\varnothing\ \text{ при }\ i\neq j;\ i,j=\overline{1,D_{l+1}},\\ \forall L\in N\ \ (((x_1,...,x_N)\in\Omega_d^L)\to(\forall\Omega_i^L,i\neq d)(\delta_i^L=0)),\end{gathered} \quad (3)$$

где $\delta_d^l(x_1,...,x_N)=\sum_{n=1}^{N}\delta_{l,n}(x_n)$ — общая поправка в зоне Ω_d^l.

С учетом разбиения пространства $X_1\times...\times X_N$ на зоны и требований (3) значение выводимой переменной можно представить в виде

$$z(x_1,...,x_N)=\lim_{L\to\infty}\sum_{l=1}^{L}\sum_{d=1}^{D(l)}p_d^l(x_1,...,x_N)\delta_d^l(x_1,...,x_N)\,, \quad (4)$$

где $p_d^l(x_1,...,x_N)=\begin{cases}1,\ (x_1,...,x_N)\in\Omega_d^l,\\ 0,\ \text{иначе}\end{cases}$ — критерий необходимости учета поправки δ_d^l в итоговом решении.

Для расчета величины δ_d^l будем использовать преобразованное соотношение Суджено (1):

$$\delta_d^l=z(x_1,...,x_L)=z_{d,0}^l+K_d^l\sum_{i=1}^{N}\left(\nu_{d,i}^l\alpha_i(x_i)\right), \quad (5)$$

где K_d^l — общий коэффициент зоны Ω_d^l; $\nu_{d,i}^l\in[0,1]$ — коэффициент влияния переменной X_i в общем значении поправки; $\alpha_i(x_i)\in[0,1]$ — значение функции принадлежности терма, расположенного в левой части продукционного правила, вычисленное на этапе фаззификации алгоритма нечеткого вывода.

Окончательно для этапа логического вывода модифицированного алгоритма Суджено значение выводимой переменной определяется как

$$z(x_1,...,x_N) = z_0 + \sum_{i=1}^{N} k_i x_i + \sum_{l=1}^{L} \frac{\sum_{d=1}^{D} p_d^l(x_1,...,x_N)\left(z_{d,0}^l + K_d^l \sum_{i=1}^{N}\left(\nu_{d,i}^l \alpha_i(x_i)\right)\right)}{\sum_{d=1}^{D} p_d^l(x_1,...,x_N)}. \quad (6)$$

Модель редукции ошибок нечеткого вывода. Для редукции ошибок решения в соответствии с выражением (6) осуществляется построение логико-лингвистической продукционной модели на основе данных о частных решениях задачи — модели редукции ошибок. Модель редукции ошибок состоит из продукционных правил четырех видов:

$$\text{если } \left(R_{x_1} = T_{l,d,k}^{R_{x_1}}\right) \text{ и ... и } \left(R_{x_N} = T_{l,d,k}^{R_{x_N}}\right), \text{ то } \left(G^k = T_{l,d}^{G^k}\right),\ k = \overline{0,N}; \quad (7)$$

$$\text{если } \left(G^0 = T_{l,d}^{G^0}\right) \text{ и ... и } \left(G^N = T_{l,d}^{G^N}\right), \text{ то } \left(\Omega = T_{l,d}^{\Omega}\right); \quad (8)$$

$$\text{если } (\Omega = T_{l,d}^{\Omega}) \text{ и } \left(D_{x_1}^l = T_{l,d}^{D_{x_1}}\right) \text{ и ... и } \left(D_{x_N}^l = T_{l,d}^{D_{x_N}}\right), \text{ то } \left(D_l = T^{D_l}\right); \quad (9)$$

$$\text{если } \left(D_1 = T^{D_1}\right) \text{ и ... и } \left(D_L = T^{D_L}\right), \text{ то } (D = D_1 + ... + D_L). \quad (10)$$

Правила (7), (8) используются для локализации области поправки, правило (9) — для вычисления величины поправки, правило (10) определяет суммарное значение поправки. Таким образом, повышение практической приемлемости решения обеспечивается не модификацией исходной экспертной модели, с помощью которой объясняется решение, а введением дополнительной модели редукции ошибок, построенной на основе обработки частных решений задачи.

Для определения параметров правил модели редукции используются алгоритмы обработки частных решений задачи, которые обеспечивают:

— разбиение пространства входных переменных $X_1 \times ... \times X_N$ на иерархию вложенных зон Ω_d^l, удовлетворяющих условию (3);

— построение продукционных правил вывода;

— определение положения функций принадлежности термов;

— расчет значений $z_{d,0}^l$, K_d^l, $\nu_{d,i}^l$ для продукционных правил вывода.

Для разбиения пространства входных переменных $X_1 \times ... \times X_N$ на зоны Ω_d^l используется алгоритм обработки точек частных решений $P_Z = \left\{\mathbf{p}_t(x_1,...,x_N,z) \mid t = \overline{1,T_Z}\right\}$, задающих значение решения z^0 при входных данных $(x_1^0,...,x_N^0)$. Зоной решения называется упорядоченная пара $\Omega = < B, \mathbf{c} >$, где $\mathbf{c} \in P_Z$ — радиус-вектор основания зоны, $B = \{\mathbf{b}_1,...,\mathbf{b}_N\}$ — система из N линейно независимых векторов (базис зоны), причем $\mathbf{b}_i = \mathbf{p}_t - \mathbf{c}$.

В процессе построения базис зоны может включать в себя как векторы стандартного базиса, так и векторы, образованные с помощью частных решений. Обработка частных решений строится таким образом, чтобы, в первую очередь, в базисах зон осуществить замещение векторов стандартного базиса векторами, образованными с помощью частных решений. При невозможности дополнения базиса формируются дополнительные зоны, смежные с первоначальными.

Основание зоны $\mathbf{c}$ и векторы ее базиса $\{\mathbf{b}_1,...,\mathbf{b}_N\}$ задают положение гиперплоскости, определяющей признак $p_d^l(x_1,...,x_N)$ учета поправки в итоговом решении и величину $\delta_d^l(x_1,...,x_N)$ этой поправки.

Значения p_d^l и δ_d^l рассчитываются следующим образом. Представив произвольный вектор $\mathbf{x} \in X_1 \times X_2 \times ... \times X_N \times Z$ в виде суммы его ортогональных составляющих $\mathbf{x}^0 \in X_1 \times X_2 \times ... \times X_N$ и $\mathbf{x}^\perp \in Z$, выразим поправку δ_d^l для зоны Ω_d^l путем разложения $\mathbf{x}$ по базису зоны Ω_d^l:

$$\delta_d^l = \mathbf{x}^\perp = \sum_{n=1}^{N} \alpha_n \mathbf{b}_n^\perp + \mathbf{c}_{l,d}^\perp, \tag{11}$$

где α_n — коэффициенты разложения вектора $\mathbf{x}$ по базису зоны Ω_d^l.

Обозначив $B_{n\times n} = (b_{i,j})$, $B_{n\times n}^{-1} = (b_{i,j}^{-1})$, выражение (11) преобразуем к виду

$$\delta_d^l (x_1, ..., x_N) = \sum_{n=1}^{N} k_{d,n}^l x_n^* + \tilde{\delta}_d^l, \tag{12}$$

где $k_{d,n}^l = b_n^\perp \sum_{j=1}^{N} b_{j,n}^{-1}$, $\tilde{\delta}_d^l = \mathbf{c}_{l,d}^\perp - \sum_{n=1}^{N} b_n^\perp c_{l,d,n} \sum_{j=1}^{N} b_{j,n}^{-1}$.

Для расчета p_d^l требуется выполнение условий

$$(\mathbf{x}^0 - \mathbf{c}_{l,d}^0) \cdot \mathbf{h}_k^0 \geq 0 \ \forall k \in \{1, 2, ..., N\}); \ (\mathbf{x}^0 - \mathbf{c}_{l,d}^0 - \mathbf{b}_1^0) \cdot \mathbf{h}_0^0 \geq 0 \ \forall j \in \{1, 2, ..., N\}, \tag{13}$$

что справедливо при

$$q^k (x_1, ..., x_N) = \sum_{i=1}^{N} x_i q_i^k - \tilde{q}^k \geq 0 \ \forall k \in \{1, 2, ..., N\} \text{ и } \delta(x_1, ..., x_N) = \left[\sum_{i=1}^{N} x_i s_i - \tilde{s} \right] \in [0, 1], \tag{14}$$

где

$$q_i^k = b_{k,i} - \sum_{j=1}^{N} \left(\sum_{m=1}^{N-1} \left(\sum_{p=1}^{N-1} g_{n,p}^k \tilde{g}_{p,m}^k \right) g_{j,m}^k \right) b_{k,j};$$

$$\tilde{q}^k = \sum_{i=1}^{N} c_i^{l,d} \left[b_{k,i} - \sum_{j=1}^{N} \left(\sum_{m=1}^{N-1} \left(\sum_{p=1}^{N-1} g_{n,p}^k \tilde{g}_{p,m}^k \right) g_{j,m}^k \right) b_{k,j} \right];$$

$\tilde{g}_{p,m}^k$ — элемент матрицы $(G_k^T G_k)^{-1}$; $s_i = \sum_{j=1}^{N} b_{j,i}^{-1}, \ i = \overline{1, N}$; $\tilde{s} = \sum_{j=1}^{N} \left(\sum_{i=1}^{N} b_{i,j}^{-1} \right) c_j^{l,d}$; $\mathbf{h}_k, k = \overline{0, N}$, — ортогональное дополнение к системе базисных векторов граней G_k, причем $\mathbf{h}_k \cdot \mathbf{b}_k \geq 0$ при $k = \overline{1, N}$ и $\mathbf{h}_0 \cdot \mathbf{b}_0 \leq 0$.

Выражения (11), (12) для расчета поправки δ_d^l и (13), (14) — для критерия p_d^l позволяют определить положение функций принадлежности термов в правилах модели редукции ошибок. В общем случае линейная комбинация $z = k_0 + \sum_{i=1}^{N} k_i x_i$ в зоне Ω с основанием $\mathbf{c} = (c_1, ..., c_{N+1})$ может быть выражена в виде

$$z = c_{N+1} + K \sum_{i=1}^{N} \left(\nu_i \alpha_i (x_i) \right),$$

где $\nu_i = \dfrac{k_i |\Omega|_i}{K}$; $K = \sqrt{\sum_{i=1}^{N} \left(k_i |\Omega|_i\right)^2}$; $|\Omega|_i$ — протяженность зоны вдоль оси, заданной ортом стандартного базиса $\mathbf{e}_i$; $\alpha_i(x_i)$ — функция принадлежности треугольного вида [7].

Таким образом, учет частных решений задачи в модели редукции ошибок обеспечивает локальную коррекцию результатов классического алгоритма нечеткого вывода Суджено и повышает практическую приемлемость решения без изменения начальной экспертной модели задачи.

Заключение. Рассмотренная модификация нечеткого вывода позволяет снизить влияние субъективного фактора, ухудшающего качество решения из-за неточностей, вносимых экспертом при описании системы. На практике в задачах управления и распознавания в области исходных данных при недостаточной информации о закономерностях работы системы известные алгоритмы нечеткого вывода приводят к ошибочным решениям. Предложенный подход позволяет добиться желаемого решения в любой области исходных данных, включая и те, где знания эксперта, выраженные в нечетком описании системы, оказываются неточными или ошибочными. Решение в этом случае достигается с помощью набора корректировочных данных. При этом корректировочные данные приводят не к модификации созданных экспертом правил или определений характеристических функций, а к дополнению существующего описания, что позволяет сохранить смысловое содержание нечеткого вывода решения в терминах, введенных экспертом.

СПИСОК ЛИТЕРАТУРЫ

1. *Wang L., Mendel J. M.* Fuzzy basis functions, universal approximation, and orthogonal least-squares learning // IEEE Transact. Neural Networks. 1992 . Vol. 3, N 5. P. 807—814.
2. *Kosko B.* Fuzzy systems as universal approximators // IEEE Transact. on Computers. 1994. Vol. 43, N 11. P. 1329—1333.
3. *Castro J. L., Delgado M.* Fuzzy systems with defuzzification are universal approximators // IEEE Transact. on System, Man, and Cybernetics. 1995. Vol. 25, N 4. P. 629—635.
4. *Tsukamoto T.* An approach to fuzzy reasoning method // Advances in Fuzzy Set Theory and Applications. Amsterdam: North-Holland Publishing Co., 1979. P. 137—149.
5. *Жиряков С. М., Майков К. А., Рогозин О. В.* Адаптация нечеткого вывода к критическим зонам ошибок управления в задачах управления // Приборы. 2009. № 2. С. 22—29.
6. *Takagi T., Sugeno M.* Fuzzy identificaton of systems and its applications to modeling and control // IEEE Transact. on System, Man, and Cybernetics. 1985. Vol. 15, N 1. P. 116—132.
7. *Круглов В. В., Дли М. И.* Нечеткая логика и искусственные нейронные сети. М.: Физматлит, 2001. 224 с.

Сведения об авторах

Владимир Анатольевич Терехов — канд. техн. наук; Московский государственный технический университет „Московский институт радиоэлектроники и автоматики“, кафедра технической электродинамики и электроники; профессор

Константин Анатольевич Майков — д-р техн. наук, профессор; Московский государственный технический университет им. Н. Э. Баумана, кафедра программного обеспечения ЭВМ и информационных технологий; E-mail: maikov@mx.bmstu.ru

Сергей Михайлович Жиряков — канд. техн. наук; ОАО «Российская самолетостроительная корпорация „МиГ“», Инженерный центр „ОКБ им. А. И. Микояна“, Москва; E-mail: zs-mailbox@mail.ru

Рекомендована Московским институтом радиоэлектроники и автоматики

Поступила в редакцию 28.06.12 г.