О качестве классификации объектов на основе нечетких правил

А. П. Рыжов

Введение

Достаточно очевидно, что качество классификации объектов на основе системы нечетких правил зависит как от качества описания объектов, так и от качества самой системы нечетких правил. Формализацией качества описания объектов является понятие степени нечеткости [1, 6] . Аналогичным образом вводятся понятия качества (степени нечеткости) системы нечетких правил. Изучается их связь в частных случаях. Приводятся результаты компьютерного моделирования для общего случая.

1. Постановка задачи

Пусть объект описывается конечным набором признаков $A = \{A_1, \ldots, A_n\}$. Каждому признаку A_i ставится в соответствие множество U_i его «физических» значений и множество $\{a_{1i}, \ldots, a_{n_i i}\}$ лингвистических значений $(1 \leqslant i \leqslant n)$. Каждому такому лингвистическому значению a_{ji} ставится в соответствие функция принадлежности $\mu_{a_{ji}}(u_i)$ в универсальном множестве U_i $(1 \leqslant j \leqslant n_i)$.

Пусть, кроме этого, есть K классов C_k, $k \in \{1 \ldots K\}$. Информация о классах в терминологии [1] задана в виде совокупности r правил:

$$\left.\begin{array}{l}
\text{Если } A_1 = a^1_{i_1 1} \text{ и } A_2 = a^1_{i_2 2} \text{ и } \ldots \text{ и } A_n = a^1_{i_n n}, \\
\qquad\qquad\qquad\qquad \text{то } \mu_{C_1}(O) = \eta^1_1 \text{ и } \ldots \text{ и } \mu_{C_k}(O) = \eta^1_K \\
\text{Если } A_1 = a^2_{i_1 1} \text{ и } A_2 = a^2_{i_2 2} \text{ и } \ldots \text{ и } A_n = a^2_{i_n n}, \\
\qquad\qquad\qquad\qquad \text{то } \mu_{C_1}(O) = \eta^2_1 \text{ и } \ldots \text{ и } \mu_{C_k}(O) = \eta^2_K \\
\qquad\qquad \vdots \qquad\qquad\qquad\qquad \vdots \\
\text{Если } A_1 = a^r_{i_1 1} \text{ и } A_2 = a^r_{i_2 2} \text{ и } \ldots \text{ и } A_n = a^r_{i_n n}, \\
\qquad\qquad\qquad\qquad \text{то } \mu_{C_1}(O) = \eta^r_1 \text{ и } \ldots \text{ и } \mu_{C_k}(O) = \eta^r_K
\end{array}\right\} \quad (1)$$

где объект O (правая часть правил) имеет значения признаков, перечисленных в левой части правил. Обозначим эту совокупность правил через F.

Будем называть систему (1) нечетким классификатором.

На вход классификатора подается вектор вещественных чисел $u^* = (u_1, \ldots, u_n)$ («физические» значения признаков объекта). Выходом является вектор $M^* = (M_1, \ldots, M_K)$, описывающий принадлежность объекта классам $C_1, \ldots, C_K$.

Нас интересуют ответы на следующие вопросы:

1) Можно ли определить качество системы правил F, отражающее совокупную внутреннюю неопределенность входящих в нее правил?
2) Как связаны качество системы правил F с качеством классификации?

2. Необходимые определения

Для ответа на сформулированные вопросы нам понадобится формализовано описывать множества значений качественных признаков. Моделью таких структур является семантическое пространство [6], которое можно рассматривать как частный случай понятия лингвистической переменной, введенной Заде в [2]. Напомним основные определения. Рассмотрим t нечетких переменных с именами $a_1, a_2, \ldots, a_t$, заданных на одном универсальном множестве. Будем называть такую совокупность *семантическим пространством* s_t. Введем систему ограничений для функций принадлежности нечетких переменных, составляющих s_t. Будем для простоты обозначать функцию принадлежности a_j через μ_j. Будем считать, что:

1) $\forall \mu_j$ $(1 \leqslant j \leqslant t)$ $\exists U_j^1 \neq \varnothing$, где $U_j^1 = \{u \in U : \ \mu_j = 1\}$, U_j^1 есть отрезок или точка;
2) $\forall j$ $(1 \leqslant j \leqslant t)$ μ_j не убывает слева от U_j^1 и не возрастает справа от U_j^1 (так как, согласно 1, U_j^1 является отрезком или точкой, понятия «слева» и «справа» определяются однозначно);
3) $\forall j$ $(1 \leqslant j \leqslant t)$ μ_j имеет не более двух точек разрыва первого рода.

Будем для простоты обозначать требования 1–3 через L. Введем также систему ограничений для совокупностей функций принадлежности нечетких переменных, образующих s_t. А именно, будем считать, что:

4) $\forall u \in U$ $\exists j$ $(1 \leqslant j \leqslant t)$: $\mu_j(u) > 0$;
5) $\forall u \in U$ $\sum\limits_{j=1}^{t} \mu_j(u) = 1$.

Будем для простоты обозначать требования 4, 5 через G. Будем называть семантическое пространство, состоящее из нечетких переменных, функции принадлежности которых удовлетворяют требованиям 1–3, а их совокупности — требованиям 4 и 5, *полным ортогональным семантическим пространством* и обозначать его $G(L)$. Достаточно подробная интерпретация требований 1–5 приведена в [4, 6].

Под степенью нечеткости $s_t \in G(L)$ будем понимать значение функционала $\xi(s_t)$, определенного на элементах $G(L)$ и принимающего значения в $[0, 1]$ (то есть $\xi : \ G(L) \to [0, 1]$), удовлетворяющего следующим условиям (аксиомам):

A1. $\xi(s_t) = 0$, если s_t представляет собой совокупность характеристических функций;

A2. Пусть s_t, $s'_{t'} \in G(L)$, t и t' могут быть равны или не равны друг другу. Тогда

$$\xi(s_t) \leqslant \xi(s'_{t'}), \text{ если } d(s_t, \hat{s}_t) \leqslant d(s'_{t'}, \hat{s}'_{t'}),$$

где $\hat{s}_t$ — ближайшая к s_t совокупность характеристических функций, $d(s_t, \hat{s}_t)$ — расстояние между семантическим пространством и ближайшей совокупностью характеристических функций [6].

Смысл аксиом следующий: из двух семантических пространств то имеет меньшую неопределенность, которое более «похоже» на пространство из совокупностей характеристических функций.

Система нечетких правил (1) содержит логические связки. В теории нечетких множеств общим представлением логических связок являются треугольные нормы (обобщенное «И») и треугольные конормы (обобщенное «ИЛИ»).

Треугольной нормой (сокращенно t-нормой) называется двухместная действительная функция $\top : [0,1] \times [0,1] \to [0,1]$, удовлетворяющая следующим условиям:

- $\top(0,0) = 0$, $\top(\mu_A, 1) = \top(1, \mu_A) = \mu_A$ (ограниченность);
- $\top(\mu_A, \mu_B) = \top(\mu_B, \mu_A)$ (коммутативность);
- $\top(\mu_A, \top(\mu_B, \mu_C)) = \top(\top(\mu_A, \mu_B), \mu_C)$ (ассоциативность);
- $\top(\mu_A, \mu_B) \leqslant \top(\mu_C, \mu_D)$, если $\mu_A \leqslant \mu_C$, $\mu_B \leqslant \mu_D$ (монотонность).

Треугольной конормой (сокращенно t-конормой) называется двухместная действительная функция $\perp : [0,1] \times [0,1] \to [0,1]$, удовлетворяющая следующим условиям:

- $\perp(1,1) = 1$, $\perp(0, \mu_A) = \perp(\mu_A, 0) = \mu_A$ (ограниченность);
- $\perp(\mu_A, \mu_B) = \perp(\mu_B, \mu_A)$ (коммутативность);
- $\perp(\mu_A, \perp(\mu_B, \mu_C)) = \perp(\perp(\mu_A, \mu_B), \mu_C)$ (ассоциативность);
- $\perp(\mu_A, \mu_B) \geqslant \perp(\mu_C, \mu_D)$, если $\mu_A \geqslant \mu_C$, $\mu_B \geqslant \mu_D$ (монотонность).

Более подробно с примерами и свойствами t-норм и t-конорм можно ознакомиться в [3, 6].

3. Степень нечеткости системы правил логического вывода

Зафиксируем некоторый номер i $(1 \leqslant i \leqslant n)$ и рассмотрим совокупность $\{a_{1i}, \ldots, a_{n_i i}\}$ лингвистических значений признака A_i. Будем считать, что этот набор значений является полным ортогональным семантическим пространством, то есть совокупность функций принадлежности $s_{n_i} = \{\mu_{a_{1i}}, \ldots, \mu_{a_{n_i i}}\}$ удовлетворяет условиям $G(L)$. Результат классификации — вектор M^* — можно интерпретировать как совокупность k нечетких множеств, заданных на одноэлементном универсальном множестве. Для удовлетворения условиям G будем требовать выполнения ортогональности для этой совокупности (требование полноты, очевидно, выполняется). Это означает, что для всех l $(1 \leqslant l \leqslant r)$ выполнено:

$$\sum_{k=1}^{K} \eta_k^l = 1. \tag{2}$$

Введем понятие степени нечеткости правила системы F способом, аналогичным введению понятия степени нечеткости для полных ортогональных семантических пространств. Зафиксируем некоторое l $(1 \leqslant l \leqslant k)$ и рассмотрим величину

$$\zeta(F, l) = 1 - (\eta_{i_1^*}^l - \eta_{i_2^*}^l), \tag{3}$$

где

$$\eta_{i_1^*}^l = \max_{k=1,\ldots,K} \eta_k^l, \quad \eta_{i_2^*}^l = \max_{\substack{k=1,\ldots,K \\ k \neq i_1^*}} \eta_k^l. \tag{4}$$

Достаточно очевидно, что степень нечеткости правила (3) равна нулю (минимальна), когда правило однозначно относит объект к одному из классов ($\exists k$ $(1 \leqslant k \leqslant K)$: $\eta_k^l = 1$, и, как следствие ортогональности (2), $\eta_j^l = 0$ для всех $j \neq k$ $(1 \leqslant j \leqslant K)$) и равна единице (максимальна), когда правило относит объект ко всем классам с равной уверенностью ($\forall k$ $(1 \leqslant k \leqslant K)$ $\eta_k^l = \frac{1}{k}$).

Под степенью нечеткости системы F правил нечеткого вывода вида (1) естественно понимать

$$\Xi(F) = \frac{1}{r} \sum_{l=1}^{r} \zeta(F, l), \tag{5}$$

где $\zeta(F, l)$ описывается формулой (3) с учетом (4).

Достаточно очевидно, что степень нечеткости $\Xi(F) = 0$ (минимальна), если все правила в системе F являются четкими и $\Xi(F) = 1$ (максимальна), если все правила, не зависимо от левой части, относят объекты с равной уверенностью ко всем классам.

Итак, при работе нечеткого классификатора имеется два вида нечеткости: нечеткость входной информации (левая часть правил) и нечеткость системы правил логического вывода. Первая из них измеряется как степень нечеткости соответствующих полных ортогональных семантических пространств, вторая — как степень нечеткости (5).

Рассмотрим функционирование нечеткого классификатора аналогично функционированию нечеткого контроллера [3].

Итак, входом нечеткого классификатора является вектор $u^* = (u_1, \ldots, u_n)$ вещественных чисел, выходом — вектор $M^* = (M_1, \ldots, M_K)$, описывающий принадлежность объекта классам $C_1, \ldots, C_K$. Выход классификатора вычисляется следующим образом.

Вычисляются величины

$$m_l(u^*) = \mathop{\top}_{k=1}^{n} \mu^l_{i_k k}(u_k), \quad (1 \leqslant l \leqslant r), \tag{6}$$

где $\mu^l_{i_k k}(u_k)$ — степень принадлежности числа u_k множеству $a^l_{i_k k}$ в (1).

Содержательно $m_l(u^*)$ описывает, насколько вектор u^* удовлетворяет правой части l-правила.

Далее вычисляются величины

$$M^l_k(u^*) = \top(m_l(u^*), \eta^l_k), \quad (1 \leqslant l \leqslant r,\ \ 1 \leqslant k \leqslant K), \tag{7}$$

где η^l_k — степень принадлежности описания объекта к k классу в l-правиле системы нечетких правил вывода F (1); $m_l(u^*)$ описывается формулой (6).

Содержательно $M^l_k(u^*)$ описывает, с какой степенью уверенности l-правило относит вектор u^* к k классу в (13).

Далее вычисляются

$$M_k(u^*) = \mathop{\perp}_{l=1}^{r} M^l_k(u^*), \quad (1 \leqslant k \leqslant K), \tag{8}$$

где $M_k^l(u^*)$ описывается (7).

$M_k(u^*)$ описывает, с какой степенью уверенности система правил (1) относит вектор u^* к k классу.

Аналогично (3), под степенью нечеткости классификации объекта u^* системой F, будем понимать

$$\Omega(F, u^*) = 1 - (M_{i_1^*}(u^*) - M_{i_2^*}(u^*)), \tag{9}$$

где

$$M_{i_1^*}(u^*) = \max_{k=1,\dots,K} M_k(u^*), \quad M_{i_2^*}(u^*) = \max_{\substack{k=1,\dots,K \\ k \neq i_1^*}} M_k(u^*). \tag{10}$$

Так же, как и при анализе (3), мы можем заметить, что степень нечеткости (9) равна нулю (минимальна), когда F однозначно относит объект u^* к одному из классов и равна единице (максимальна), когда F относит объект u^* ко всем классам с равной уверенностью.

Аналогично степени нечеткости множества, под степенью нечеткости классификатора F будем понимать усредненную степень нечеткости классификации им объектов при условии равномерного распределения объектов в универсальном множестве $U^* = U_1 \times \dots \times U_n$:

$$\Omega(F) = \frac{1}{|U^*|} \int\limits_{U^*} \Omega(F, u^*)\, du^* = \frac{1}{|U^*|} \int\limits_{U^*} (1 - (M_{i_1^*}(u^*) - M_{i_2^*}(u^*)))\, du^*, \tag{11}$$

где $M_{i_1^*}(u^*)$, $M_{i_2^*}(u^*)$ описываются формулами (10).

Теперь мы можем ответить на основной вопрос данной работы: как зависит качество классификации объектов по их нечетким описаниям на основе системы нечетких правил от качества описания объектов и качества самой системы нечетких правил.

4. О качестве классификации объектов нечетким классификатором

Утверждение 1. *Если* $\Xi(F) = 1$, *то* $\Omega(F) = 1$.

Доказательство. Рассмотрим (11) с учетом (10) и формул (6)–(8). Подставляя их в (11), получаем следующее развернутое выражение для $\Omega(F)$:

$$\Omega(F) = \frac{1}{|U^*|}\int\limits_{U^*}\left(1 - \left(\mathop{\perp}_{l=1}^{r} \top(m_l(u^*), \eta^l_{i_1^*}) - \mathop{\perp}_{l=1}^{r} \top(m_l(u^*), \eta^l_{i_2^*})\right)\right) du^*. \tag{12}$$

Из (5) следует, что при выполнении условия утверждения $\forall l$ $(1 \leqslant l \leqslant r)$ $\zeta(F, l) = 1$. Это, в свою очередь, означает, что $\forall l$ $(1 \leqslant l \leqslant r)$ $\eta^l_{i_1^*} = \eta^l_{i_2^*} = \max\limits_{k=1,\dots,K} \eta^l_k$. Обозначим $\max\limits_{k=1,\dots,K} \eta^l_k$ через η^l. С учетом данного замечания, (12) можно переписать следующим образом:

$$\Omega(F) = \frac{1}{|U^*|}\int\limits_{U^*}\left(1 - \left(\mathop{\perp}_{l=1}^{r} \top(m_l(u^*), \eta^l) - \mathop{\perp}_{l=1}^{r} \top(m_l(u^*), \eta^l)\right)\right) du^* = 1. \tag{13}$$

Утверждение доказано.

Данное утверждение означает, что при максимальной неопределенности F неопределенность работы нечеткого классификатора будет также максимальна; она никак не зависит от степени неопределенности входной информации.

Утверждение 2. *Если* $\Xi(F) = 0$, *то*

$$\Omega(F) = \frac{1}{|U^*|}\int\limits_{U^*}\left(1 - \mathop{\perp}_{l=1}^{r} m_l(u^*)\right) du^*. \tag{14}$$

Доказательство. Из (5) следует, что при выполнении условия утверждения $\forall l$ $(1 \leqslant l \leqslant r)$ $\zeta(F, l) = 0$. Это в свою очередь означает, что $\forall l$ $(1 \leqslant l \leqslant r)$ $\eta^l_{i_1^*} = 1$ и $\eta^l_{i_2^*} = 0$. С учетом данного замечания (12) можно переписать следующим образом:

$$\Omega(F) = \frac{1}{|U^*|}\int\limits_{U^*}\left(1 - \left(\mathop{\perp}_{l=1}^{r} \top(m_l(u^*), 1) - \mathop{\perp}_{l=1}^{r} \top(m_l(u^*), 0)\right)\right) du^*. \tag{15}$$

Из свойств ограниченности t-нормы [6] непосредственно следует, что $\top(m_l(u^*), 0) = 0$ и $\top(m_l(u^*), 1) = m_l(u^*)$. Подставляя эти значения в (15), получаем доказательство утверждения.

Данное утверждение означает, что при отсутствии неопределенности в F неопределенность работы нечеткого классификатора будет определяться только тем, насколько в среднем объекты из U^* удовлетворяют самому «худшему» правилу системы F, то есть в некотором смысле адекватностью самой системы правил реальному процессу.

Непосредственно из утверждения 2 получаем следующее следствие.

Следствие 1. *Если* $\Xi(F) = 0$ *и* $\forall u^* \in U^*$ $\exists l$ *(*$1 \leqslant l \leqslant r$*):* $\forall k$ *(*$1 \leqslant k \leqslant n$*)* $\mu^l_{i_k}(u_k) = 1$, *то*

$$\Omega(F) = 0.$$

Доказательство. Для доказательства утверждения достаточно заметить, что, если выполнено $\forall k$ $(1 \leqslant k \leqslant n)$ $\mu^l_{i_k}(u_k) = 1$, то, согласно (6) и условию ограниченности t-нормы, $m_l(u^*) = 1$. Это в свою очередь означает, что $\mathop{\perp}\limits_{l=1}^{r} m_l(u^*) = 1$ (в силу условия ограниченности для t-конорм). Выполнение последнего равенства для всех $u^* \in U^*$ означает равенство нулю подынтегральной функции в (14) на всей области интегрирования, что и доказывает следствие 1.

Следствие утверждает, что при отсутствии неопределенности в F и ситуации, когда для любого объекта найдется правило, левой части которого объект полностью удовлетворяет, классификатор будет однозначно классифицировать все объекты (неопределенность классификации будет равно нулю).

Заметим, что для реализации такой ситуации необходимо, чтобы:

- для любого объекта существовало значение признака, которому он удовлетворяет со степенью принадлежности 1 (не трудно заметить, что для полных ортогональных семантических пространств это означает отсутствие нечеткости);
- система правил F должна иметь в левой части все возможные комбинации значений признаков объектов (это своеобраз-

ное требование полноты для F; не трудно подсчитать, что в этом случае F должна содержать $N_1 \times N_2 \times \cdots \times N_{N_0}$ правил).

Таким образом, равенство нулю степени нечеткости описаний объектов и степени нечеткости системы F являются необходимыми (но не достаточными) условиями равенства нулю степени нечеткости классификации объектов нечетким классификатором.

В заключение отметим, что приведенные здесь утверждения описывают общие свойства нечеткого классификатора. Для решения конкретной задачи, естественно, можно добиться более лучших результатов выбором различных реализаций t-норм и t-конорм на основе генетических алгоритмов или нейронных сетей [5]. В качестве примера можно привести следующую иллюстрацию.

Для модельной задачи (два признака — вес и размер; три класса — удобный для переноски объект, удовлетворительный для переноски объект, неудобный для переноски объект) были реализованы два алгоритма:

- запрограммирована работа нечеткого классификатора с t-нормой min и t-конормой max;
- для той же системы нечетких правил была запрограммирована и обучена нейронная сеть.

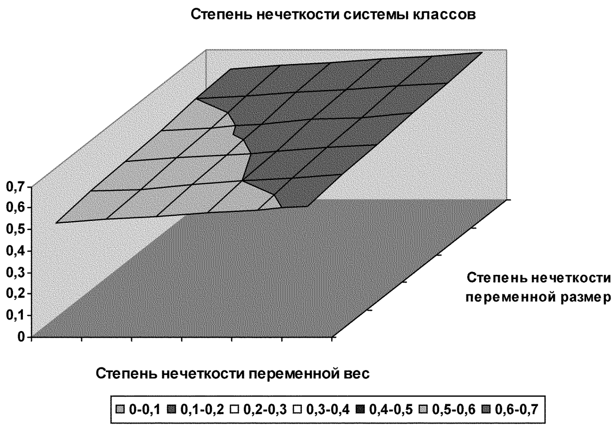

Рис. 1.

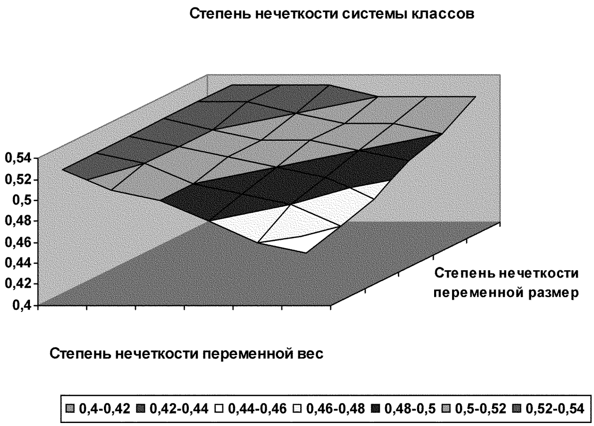

Рис. 2.

Для достаточно «мелкой» сетки был вычислен выход как в первом, так и во втором случаях. В результате такого численного эксперимента было получено приближение $\Omega(F)$ (11) и усредненная по всем объектам степень нечеткости классификации объектов нейронной сетью. Результаты одного из таких экспериментов представлены на рис. 1 и 2 соответственно. Сравнивая эти рисунки, нетрудно заметить, что степень нечеткости системы классов, полученной в результате работы нейронной сети, меньше либо равна степени нечеткости системы классов, полученной в результате работы нечеткого классификатора. Таким образом, нейронная сеть реализует более «четко» работающий классификатор. Это связано с тем, что нейронная сеть находит более адекватные формализации t-норм и t-конорм для данной конкретной задачи.

Список литературы

[1] Журавлев Ю. И. Об алгебраическом подходе к решению задач распознавания и классификации // Проблемы кибернетики. 1978. Вып. 33. С. 28–57.

[2] Заде Л. А. Понятие лингвистической переменной и его применение к принятию приблизительных решений. М.: Мир, 1976. 165 с.

[3] Нечеткие множества в моделях управления и искусственного интеллекта. Под ред. Поспелова Д. А. М.: Наука, 1986. 395 с.

[4] Рыжов А. П. Оценка степени нечеткости и ее применение в системах искусственного интеллекта // Интеллектуальные системы. Т. 1, вып. 1–4. М., 1996. С. 205–216.

[5] Рыжов А. П., Федорова М. С. Генетические алгоритмы в задаче выбора операторов агрегирования информации в системах информационного мониторинга // V Всероссийская конференция «Нейрокомпьютеры и их применение». Сборник докладов. Москва, 17–19 февраля 1999 г. С. 267–270.

[6] Рыжов А. П. Элементы теории нечетких множеств и измерения нечеткости. М.: Диалог-МГУ, 1998. 116 с.