Балашов О.В., Букачев Д.С. Подход к определению качественных характеристик объектов // Международный журнал информационных технологий и энергоэффективности. – 2021. – Т.6 №4 (22). – с. 18-23

ОТКРЫТАЯ НАУКА
издательство

Международный журнал информационных технологий и энергоэффективности

Сайт журнала:
http://www.openaccessscience.ru/index.php/ijcse/

УДК 004.05

ПОДХОД К ОПРЕДЕЛЕНИЮ КАЧЕСТВЕННЫХ ХАРАКТЕРИСТИК ОБЪЕКТОВ

[1]Балашов О.В., [2]Букачев Д.С.
[1]Смоленский филиал АО «Радиозавод», Россия, (214027, г. Смоленск, улица Котовского, 2), e-mail: smradio@mail.ru
[2]ФГБОУ ВО Смоленский государственный университет, Смоленск, Россия (214000, г. Смоленск, ул. Пржевальского, 4), e-mail: dsbuka@yandex.ru

В статье рассмотрен способ определения нечёткости логико-лингвистической шкалы при определении качественных характеристик объектов управления и предложена методика выбора оптимального множества значений качественного признака.

Ключевые слова: информация, неопределённость, выбор, функция принадлежности, отношение предпочтения.

APPROACH TO DETERMINING THE QUALITATIVE CHARACTERISTICS OF OBJECTS

[1]Balashov O.V., [2]Bukachev D.S.
[1]Smolensk branch of joint-stock company "Radio factory", Russia, (214027, Smolensk, street Kotovskogo, 2), e-mail: smradio@mail.ru
[2]Federal State Educational Institution of Higher Education Smolensk State University, Smolensk, Russia (214000, Smolensk, street Przewalski, 4), e-mail: dsbuka@yandex.ru

The article discusses a method for determining the fuzziness of the logical-linguistic scale when determining the qualitative characteristics of control objects and suggests a method for choosing the optimal set of values of a qualitative attribute.

Keywords: information, indeterminacy, choice, membership function, preference relationship.

В настоящее время значительное число компаний заказывают и внедряют в информационные системы в качестве приложений системы поддержки принятия решений (СППР) руководителей различных уровней. Разработка программного обеспечения таких СППР требует специфических подходов к их проектированию, особенно к проектированию механизмов обработки информации. Опыт работ по созданию СППР показывает, что существует проблема формализации информации и приведения её к виду, удобному для машинной обработки. При создании баз данных систем поддержки принятия решений, а точнее в процессе разработки логико-лингвистических шкал, решается задача выбора того или иного лингвистического значения из заданного множества при описании характеристик реальных объектов путём обработки экспертной информации. Решая данную задачу, эксперты испытывают определенные затруднения. Перед проектировщиками встаёт задача: исходя из анализа структуры множества шкальных значений нечёткой лингвистической шкалы (НЛШ), разработать правило, используя которое эксперт оценивал бы объекты с минимальными трудностями.

Процесс оценки свойств объектов можно рассматривать как некоторую процедуру измерения, поэтому рассмотрим ряд основных понятий теории измерений [1].

Пусть Q – множество объектов с определённым на нем набором отношений R_i, $(i \in I)$. Пара $\langle Q;\ R_i,\ i \in I \rangle$ определяется как система с отношениями. Если Q интерпретируется как множество объектов реального мира, то система с отношениями (СО) называется эмпирической, если под Q понимается числовая ось – числовой СО [1].

Под шкалой понимается гомоморфизм l эмпирической СО $\langle Q;\ V_i,\ i \in I \rangle$ в числовую СО $\langle A;\ W_i,\ i \in I \rangle$. Элементы множества A при этом называются множеством шкальных значений шкалы l. Более строго, под нечёткой лингвистической шкалой понимается гомоморфизм l эмпирической СО $\langle Q;\ V_i,\ i \in I \rangle$ в СО $\langle \tilde{R};\ W_i,\ i \in I \rangle$, где $\tilde{R}$ – совокупность поименованных нечётких чисел, заданных на R_i; СО $\langle \tilde{R};\ W_i,\ i \in I \rangle$ называется лингвистической числовой СО [2]. Пример множества шкальных значений НЛШ «Рост» достаточно подробно приведен в [3].

Интегральная характеристика НЛШ – степень нечеткости, её содержательный смысл – это степень сомнений (колебаний) эксперта, которые он испытывает при описании объектов в данной шкале. Считается, что для каждой точки опорного пространства существует хотя бы одно шкальное значение, имеющее в этой точке ненулевое значение функции принадлежности, то есть

$$\forall u \in U \ \ \exists i\ (1 < i < t) : \mu_{ai} > 0. \tag{1}$$

Это ограничение не является искусственным, так как, если оно не выполняется, то множество точек $U' = \{u \in U' : \forall i\ (i = 1,\ \ldots,\ t)\ \mu_{ai}\ (u) = a\}$ можно без ущерба удалить из U, то есть в качестве универсума рассматривать множество $U \backslash U'$. Это говорит о том, что ни одному реальному объекту в точках множества U не соответствует ни одно шкальное значение, то есть шкала плохо определена.

Под степенью нечёткости НЛШ понимается степень нечёткости множества $\tilde{R}$ её шкальных значений. Для определения степени нечёткости лингвистической шкалы l_t (t – число шкальных значений), вводится понятие степени её нечёткости) в точке $u \in U$ $(\eta(l_t\,,\ u))$.

На содержательном уровне под $\eta(l_t,\ u)$ понимается степень сомнения (колебаний) эксперта в выборе того или иного шкального значения данной шкалы в рассматриваемой точке опорного пространства при описании реальных объектов. Очевидно, величина этих колебаний обратно пропорциональна разности между максимальным и ближайшим к нему значениями функции принадлежности $\mu_{a1}(u)$, $\ldots \mu_{at}(u)$, шкальных значений.

Для иллюстрации приведем следующий пример. Пусть на складе имеется запас ресурса с максимальным значением в 48 единиц. При оценке запаса u_1 = 20 единиц и u_2 = 45 единиц эксперты практически без колебаний выбирают одно из шкальных значений «средний» и «большой» соответственно. При описании запаса ресурса u_3 = 30 единиц эксперт начинает колебаться в выборе одного из названных шкальных значений, и эти колебания становятся максимальными в точке u_4 = 35 единиц. Таким образом,

$$\eta(l_t, u) = 1 - (\mu_{ai_1}(u) - \mu_{ai_2}(u)), \tag{2}$$

где

$$\mu_{ai_1}(u) = \max_{1 \le j \le t} \mu_{a_j}(u), \tag{3}$$

$$\mu_{ai_2}(u) = \max_{1 \le j \le t;\, j \ne i_1} \mu_{aj}(u). \tag{4}$$

Из формул (2) – (4) следует, что

$$\eta\,(l_t, u) = 1 \Leftrightarrow \forall j\ (1 \le j \le t, j \ne i)\ \exists i\ (1 \le i \le t) : \mu_{ai}(u) = 1,\ \mu_{aj}(u) = 0),$$

$$\eta\,(l_t, u) = 0 \Leftrightarrow \exists i_1, i_2\ (1 \le i_1, i_2 \le \mathrm{t}) :\ \mu_{ai_1}(u) = \mu_{ai_2}(u) = \max_{1 \le j \le t} \mu_{a_j}(u).$$

Возвращаясь к примеру ($l_t = l_3$= «Запас ресурса»), получается:

$$\eta\,(l_3, u_2) = 0;\ \eta\,(l_3, u_3) = 0{,}5;\ \eta(l_t, u_4) = 1.$$

Таким образом, $\eta(l_t, u)$ отражает степень колебаний эксперта при выборе того или иного шкального значения в точке $u \in U$.

Необходимо отметить, что такое определение степени нечёткости лингвистической шкалы в точке универсума (формулы (2) – (4)), не учитывают влияния оставшихся (t–2) шкальных значений в данной точке и наиболее адекватно отражают случай пересечения функций принадлежности двух из них.

В силу ряда субъективных причин (ограниченного числа экспертов, различий в уровне их квалификации, неудачном выборе множества шкальных значений и др.) может наблюдаться ситуация плохого определения лингвистической шкалы l_t, когда в некоторой точке $u \in U$ могут пересекаться функции принадлежности трех, четырех и больше (вплоть до t) шкальных значений. Можно пойти двумя путями: усложнять формулы (2) – (4) или рассматривать некоторое подмножество всех шкал – G-шкалы, определяемые следующим образом:

$$l_t \in F \Leftrightarrow \forall u \in U:$$

либо $\exists\ \mu_{ai_1}(u),\ \mu_{ai_2}(u),\ \mu_{ai_3}(u)\,(1 \le i_1,\ i_2,\ i_3 \le t) :\ \mu_{ai_1}(u) \ne 0,\ \mu_{ai_2}(u) \ne 0,\ \mu_{ai_3}(u) \ne 0;$

либо $\forall \mu_{ai}(u)\ (\ \mu_{ai_1}(u) > \mu_{ai_2}(u) > \ldots\ \mu_{ai_k}(u) > 0,\ k \ge 3,\ 1 \le i_1, \ldots, i_k \le t)$:

$$\mu_{ai_1}(u) - \mu_{ai_3}(u) > c(\mu_{ai_1}(u) - \mu_{ai_2}(u)),\ c = const,\ c > 2.$$

G-шкалы не только удобны для теоретического анализа, но и наиболее часто встречаются на практике, так как описанные выше требования на содержательном уровне означают необходимость того, чтобы используемые понятия (шкальные значения) достаточно различались между собой семантически, не описывали одни и те же объекты, не являлись омонимами.

Используя понятие степени нечёткости шкалы в точке $u \in U$, определяется степень нечёткости $\nu(l_t)$ лингвистической шкалы как средняя степень её нечёткости во всем множестве U:

$$\nu(l_t) = \frac{1}{|U|}\int_U \eta(l_t, u)du, \tag{5}$$

где $|U|$ – мощность универсума U, $\eta(l_t, u)$ определяемая формулами (2) – (4).

Очевидно, что:

$\forall l_t$: $\nu(l_t) = 0 \Leftrightarrow \forall u \in U\ \forall j\ (1 \le j \le t)\ \exists i_1^*\ (1 \le i_1^* \le t,\ j \ne i_1^*)$: $\mu_{ai_1}(u) = 1,\ \mu_{aj}(u) = 0$ (то есть в абсолютно чётком случае);

$\forall l_t$: $\nu(l_t) = 1 \Leftrightarrow \forall u \in U\ \exists i_1^*,\ i_2^*,\ (1 \le i_1^*,\ i_2^*, \le \text{t})$: $\mu_{ai_1^*}(u) = \mu_{ai_2^*}(u) = \max\limits_{1 \le j \le t} \mu_{aj}(u)$.

С помощью введенного понятия степени нечёткости НЛШ можно определить степень нечёткости множества как частный случай $\nu(l_t)$.

Совокупность результатов оценок некоторого объекта в НЛШ $l_{t_1}^1, \ldots, l_{t_k}^k$ называются его лингвистическим описанием по признакам с номерами $1, \ldots, k$ [5]. Множество лингвистических описаний объектов интерпретируется как база данных информационной системы СППР. Рассмотрим показатели качества работы информационной системы – это средние индивидуальные потери информации и шумы, возникающие при поиске информации в ней, под которыми понимается следующее.

При общении с СППР лицо, принимающее решения (ЛПР) формирует запрос (например, «Выдать описания всех объектов, имеющих значение характеристики «Запас ресурса», равное «НИЗКИЙ») и получает из базы данных СППР некоторое количество описаний объектов, удовлетворяющих поисковому предписанию. При этом, если бы ЛПР знал реальные значения характеристик, выданных на запрос объектов и объектов, описания которых хранятся в базе данных, он, возможно, забраковал бы часть выданных объектов (информационный шум), а часть объектов, описания которых не были выданы на запрос, наоборот, принял бы (потери информации). Механизм возникновения таких потерь и шумов связан с размытостью элементов шкалы, с тем, что источник информации и ЛПР могли бы выбрать для описания одного и того же реального объекта разные шкальные значения НЛШ. Такие потери информации и шумы можно назвать индивидуальными. Обозначим через $B(l_t)$ и $S(l_t)$ объемы средних индивидуальных потерь информации и шумов, возникающих при поиске информации по признаку с множеством значений, состоящим из имен шкальных значений НЛШ l_t.

Для подсчета объема средних потерь информации и шумов вводится упрощение: считается, что описание $I(Q)$ реального объекта Q, хранящегося в СППР, содержит только одну поисковую характеристику, имеющую t значений a_1, a_2, $\ldots a_t$, представляющих собой множество имен шкальных значений лингвистической шкалы l_t. Для функций принадлежности, входящих в набор множества шкальных значений, кроме условия (1) выполняется требование ортогональности [5]:

$$\forall u \in U\ \sum_{j=1}^{t} \mu_{aj}(u) = 1 \tag{6}$$

Это предположение связано с выбором той или иной интерпретации понятия «функция принадлежности». По приведенной в [5] классификации такое требование определяет так называемую вероятностную интерпретацию, что, однако, не сводит данное понятие ни к функции распределения, ни к плотности вероятности [5, 6]. Выдвинутое требование выполняется или нет в зависимости от метода построения функции принадлежности, в частности, при следующем определении:

$$\mu_{aj}(u) = \frac{n_1}{n_1 + n_2},$$

где n_1 – число экспертов, относящих u к множеству a_j;

$n_1 + n_2$ – общее число экспертов.

При опросе возможны только ответы вида «да, $u \in a_j$» или «нет, $u \notin a_j$». Приведенные рассуждения позволяют сделать вывод о том, что ограничение (6) не является очень искусственным.

При проектировании СППР, для оценки объёма средних потерь информации и шумов в ходе работы с экспертами была сформулирована и доказана следующая теорема.

Теорема. Пусть l_t – некоторая лингвистическая шкала, имеющая t шкальных значений, функции принадлежности которых удовлетворяют условиям (1), (6). $\nu(l_t)$ – степень нечеткости, $B_x(l_t)$ и $S_x(l_t)$ – средние индивидуальные потери информации и шумы, возникающие при поиске информации по признаку с множеством значений X, совпадающим с множеством имен шкальных значений l_t; U – универсум.

$N(u)$ – число объектов, описания которых хранятся в базе данных СППР, имеющих фактическое значение характеристики, равное u – есть константа. Для ЛПР значения признака представляют одинаковый интерес, то есть вероятности запросов по каждому значению признака равны. Тогда

$$B_x(l_t) = S_x(l_t) = \frac{2N}{3t}\,\nu(l_t),\ N = const.$$

Кратко необходимо пояснить, что при её доказательстве использовалось (L–R)-представление нечётких множеств и формулы аналитической геометрии. При выполнении условий теоремы

$$\nu(l_t) = \frac{1}{|U|}\frac{1}{2}\sum_{j=1}^{t} d_j;\ B_x(U) = S_x(U) = \frac{N}{3t}\sum_{j=1}^{t} d_j.$$

Отсюда следует, что уменьшение степени нечёткости на $d\%$ при фиксированном числе значений характеристики ведёт к такому же сокращению средних индивидуальных потерь информации и шумов, возникающих при работе с данной характеристикой. Уменьшение же степени нечёткости при одновременном увеличении числа значений характеристики делает эту зависимость ещё более сильной.

Таким образом, исходя из вышеизложенного, можно предложить следующую методику выбора множества значений качественного признака:

- сформировать все возможные множества значений признака;

- каждое множество значений признака представить в виде множества шкальных значений лингвистической шкалы;
- для каждого множества значений вычислить степень нечёткости признака по формулам (2) – (5);
- в качестве оптимального множества значений, минимизирующего неопределённость при описании объектов, выбрать то множество, для которого степень нечёткости минимальна;
- в качестве оптимального множества значений, повышающего качество поиска информации, выбрать то множество, для которого отношение степени нечёткости к числу значений признака минимально.

Список литературы

1. Ларичев О.И. Теория и методы принятия решений. – М.: Логос, 2002.
2. Борисов А. Н., Алексеев А.В., Крумберг О. А. и др. Модели принятия решений на основе лингвистической переменной. Рига, Зинатне, 1982.
3. Заде Л. Понятие лингвистической переменной и его применение к принятию приближенных решений. М. Мир, 1976.
4. Балашов О.В., Букачев Д.С. Кондратова Н.В. Способы формализации задачи принятия решений для проектирования систем поддержки принятия решений // Международный журнал информационных технологий и энергоэффективности. – 2018. – Т.3, №1(7). – С. 25-34.
5. Аверкин А.Н., Батыршин И.З., Блишун А.Ф. и др. Нечёткие множества в моделях управления и искусственного интеллекта/ Под ред. Поспелова Д.А. М. Наука. Гл.ред.физ.-мат. лит., 1986.
6. Борисов А.Н., Алексеев А.В., Меркурьева Г.В. и др. Обработка нечёткой информации в системах принятия решений. М. Радио и связь, 1989.

References

1. Larichev O.I. Teoriya i metody prinyatiya reshenij. – M.: Logos, 2002.
2. Borisov A. N., Alekseev A.V., Krumberg O. A. i dr. Modeli prinyatiya reshenij na osnove lingvisticheskoj peremennoj. Riga, Zinatne, 1982.
3. Zade L. Ponyatie lingvisticheskoj peremennoj i ego primenenie k prinyatiyu priblizhennyh reshenij. M. Mir, 1976.
4. Balashov O.V., Bukachev D.S. Kondratova N.V. Sposoby formalizacii zadachi prinyatiya reshenij dlya proektirovaniya sistem podderzhki prinyatiya reshenij // Mezhdunarodnyj zhurnal informacionnyh tekhnologij i energoeffektivnosti. – 2018. – T.3, №1(7). – pp. 25-34.
5. Averkin A.N., Batyrshin I.Z., Blishun A.F. i dr. Nechyotkie mnozhestva v modelyah upravleniya i iskusstvennogo intellekta/ Pod red. Pospelova D.A. M. Nauka. Gl.red.fiz.-mat. lit., 1986.
6. Borisov A.N., Alekseev A.V., Merkur'eva G.V. i dr. Obrabotka nechyotkoj informacii v sistemah prinyatiya reshenij. M. Radio i svyaz', 1989.