

УДК 004.048

П. В. Пашко, доктор экономических наук
Б. И. Мороз, доктор технических наук,
декан факультета информационных систем
и технологий Академии таможенной
службы Украины
Ю. В. Ульяновская, кандидат технических
наук, доцент кафедры информационных си-
стем и технологий Академии таможенной
службы Украины
Н. В. Билык, магистр
Академии таможенной службы Украины

АНАЛИЗ ПРОБЛЕМЫ ОБРАБОТКИ НЕЧЕТКОЙ ЭКСПЕРТНОЙ ИНФОРМАЦИИ В ЗАДАЧАХ КЛАССИФИКАЦИИ ОБЪЕКТОВ

У статті розглядаються питання обробки експертних оцінок об'єктів. Проведено аналіз методів обробки експертної інформації в разі якісного і кількісного характеру даних.

В статье рассматриваются вопросы обработки экспертных оценок объектов. Проведен анализ методов обработки экспертной информации при качественном и количественном характере данных.

In the paper the problems of processing expert estimations of objects are considered. The analysis of methods of processing expert information is carried out under qualitative the and quantitative characters of the data.

Ключевые слова. Экспертные оценки, классификация, нечеткие данные.

Вступление. На качество работы большинства систем обработки данных большое влияние оказывает характер входных данных. Особое значение это имеет для информационных систем обработки данных, а также интеллектуальных систем, таких как экспертные системы, или системы поддержки принятия решения.

Задача обработки входных данных может быть разбита на задачу получения и задачу обработки экспертной информации. Процесс получения знаний от экспертов включает несколько этапов: подбор экспертов, опрос экспертов, обработка экспертных оценок. Экспертную оценку объектов можно рассматривать как процесс получения экспертной информации с помощью измерений. Сами объекты при этом могут быть описаны как с помощью качественных, так и количественных данных. Для принятия решений в системах искусственного интеллекта важно не только сравнивать объекты между собой, но и знать, во сколько раз и на сколько условных единиц один объект предпочтительнее другого. В процессе решения указанной задачи эксперты определяют признаки и показатели для описания свойств объектов, их взаимосвязей, измерения их характеристик (значений признаков и показателей). При этом большое значение имеет выбор системы критериев для оценки информационных характеристик признаков и шкалы их описания. Наиболее изученными являются проблемы экспертного ранжирования. Однако существует ряд нерешенных задач, исследованию которых посвящены работы Ю. Н. Тюрина, А. И. Орлова и других ученых [1].

Постановка задачи. Методы, используемые в настоящее время для получения экспертных оценок, достаточно многочисленны и разнообразны. Выбор метода определяется характером анализируемой информации. Для получения качественных оценок используются

© П. В. Пашко, Б. И. Мороз, Ю. В. Ульяновская, Н. В. Билык, 2011

парные сравнения, множественные сравнения, методы ранжирования и т. д. Для получения количественных оценок используются непосредственная численная оценка альтернатив, метод Черчмена – Акофа и др. [2]. Применение методов анализа и обработки экспертной информации зависит от характера измерений. Некорректное использование результатов экспертного оценивания может привести к ошибочным выводам. Избежать ошибок можно, лишь исследовав характер измерений и обусловленные им возможные методы преобразования полученной экспертной информации. В работах Л. Д. Мешалкина, А. И. Орлова, В. Б. Кузьмина исследуется проблема допустимости использования различных результатов экспертного ранжирования [1].

Результаты исследования. Сформулированные выше задачи возникают при построении информационных систем, работающих с нечеткими и неполными данными. Важное практическое значение имеет построение экспертной системы для идентификации объектов таможенного контроля, в частности определения культурной или исторической ценности предметов искусства [3]. Разработка такой системы направлена на сохранение культурного и исторического наследия нашей страны. В качестве альтернатив при формировании базы знаний экспертной системы выступают признаки объектов, подлежащих идентификации, при этом предпочтительность альтернатив определяется в зависимости от влияния значения признака на ценность объекта исследования. Для улучшения качества идентификации объекты предметной области необходимо проклассифицировать. Классификация может осуществляться несколькими способами. В первом случае критерием классификации выступают физические признаки объектов исследования. Это позволяет разбить всю предметную область на типы предметов искусства, такие как картины, иконы, музыкальные инструменты и др. Для второго типа классификации параметром классификации может выступать ценность объекта. В этом случае возникает задача разбиения всего множества признаков на классы эквивалентности и ранжирования признаков внутри каждого класса. Необходимость учитывать фактор несогласованности экспертов порождает задачу выбора адекватной шкалы для их измерения и преобразования.

Рассмотрим случай, когда экспертная информация носит количественный характер. В работе [4] для различных типов шкал получена групповая количественная оценка. Пусть заданы количественные оценки n объектов m экспертами $\vec{x}_1, \vec{x}_2, \dots, \vec{x}_n \in R^m$, где R^m – пространство оценок экспертов; $\vec{x}_k = (x_k^1, x_k^2, \dots, x_k^m)$, $k = \overline{1, n}$ – вектор-оценка k -го объекта m экспертами. Измерения считаются полностью согласованными, если они совпадают с точностью до допустимых преобразований. Это означает, что векторы $\vec{x}_1, \vec{x}_2, \dots, \vec{x}_n \in R^m$ лежат на допустимой прямой l , которая имеет следующий вид:

- для абсолютной шкалы $l = l(\vec{1}, \vec{0}) = \{ \vec{y} \in R^m \mid \vec{y} = t\vec{1}; \vec{1} \in R^m, t \in R \}$;
- для шкалы разностей $l = l(\vec{1}, \vec{b}) = \{ \vec{y} \in R^m \mid \vec{y} = t\vec{1} + \vec{b}; \vec{1}, \vec{b} \in R^m, t \in R \}$;
- для шкалы отношений $l = l(\vec{a}, \vec{0}) = \{ \vec{y} \in R^m \mid \vec{y} = t\vec{a}; \vec{a} \in R^m, t \in R \}$;
- для шкалы интервалов $l = l(\vec{a}, \vec{b}) = \{ \vec{y} \in R^m \mid \vec{y} = t\vec{a} + \vec{b}; \vec{a}, \vec{b} \in R^m, t \in R \}$.

Мерой близости S^2 при введенных понятиях считается минимум суммы квадратов расстояний от $\vec{x}_i, i = \overline{1, n}$ до регрессионной прямой l , где минимум берется по всем допустимым прямым:

$$S^2 = \min_l \sum_{i=1}^n d^2(\vec{x}_i, l) = \min_l \sum_{i=1}^n \min_{y \in l} \|\vec{x}_i - \vec{y}\|^2. \quad (1)$$

Результатом экспертизы в соответствии с работой является вектор $\vec{T} = \{t_1, t_2, \dots, t_n\}$, где t_i – координаты точек проекций вектор-оценок $\vec{x}_i, i = 1, \dots, n$ на прямую l , которая выбирается в соответствии с типом шкалы, в которой производились измерения.

Для измерений, проведенных в абсолютной шкале, групповой оценкой является среднее арифметическое оценок экспертов: $S^2 = \sum_{i=1}^n \|\vec{x}_i - x_i^* \vec{1}\|^2, i = \overline{1, n}$. Для шкалы разностей допустимым преобразованием является преобразование сдвига $f(t) = \vec{t}\vec{1} + \vec{b}; \vec{1}, \vec{b} \in R_m$, то есть меняется лишь начало отсчета. Мера близости $S^2 = \min_l \sum_{i=1}^n \min_{y \in l} \|\vec{x}_i - \vec{y}\|^2 = \min_{\vec{b} \in R^m} \sum_{i=1}^n \min_{t \in R} \|\vec{x}_i - \vec{b} - t\vec{1}\|^2$. Для шкалы отношений $S^2 = \min_{a \in R^m} \sum_{i=1}^n \min_{t \in R} \|\vec{x}_i - t\vec{a}\|^2$.

Рассмотрим случай, когда экспертная информация носит качественный характер [5]. На сегодняшний день разработано достаточно много методов получения знаний от эксперта и обработки этих знаний с целью построения функций принадлежности. Все методы можно разделить на прямые и косвенные, а также на методы для одного эксперта и методы для группы экспертов. Для получения исходных данных используется прямой метод для группы экспертов, который состоит в следующем. Пусть есть n экспертов, часть из которых на вопрос о принадлежности элемента $x \in X$ нечеткому множеству \check{A} отвечает положительно. Обозначим их число через n_1 . Вторая часть экспертов, а именно $n_2 = n - n_1$, отвечает на этот вопрос отрицательно. Тогда принимается, что $\mu_{\check{A}}(x) = n_1/n$. В описанной процедуре фаззификации для полного задания лингвистической переменной необходимо построение функции принадлежности $\mu_{\check{A}}(x)$ всех нечетких множеств, которые задают и другие нечеткие значения из терм – множества лингвистической переменной. Степень принадлежности имеет в результате вероятностную интерпретацию. Функция принадлежности должна быть нормализована, то есть верхняя граница функции принадлежности должна быть равна единице: $\sup_{x \in X} \mu_{\check{A}}(x) = 1$. Второй метод базируется на парных сравнениях, выполняемых одним экспертом. Исходной информацией для построения функций принадлежности являются экспертные парные сравнения. Для каждой пары элементов x_i, x_j универсального множества эксперт оценивает преимущество одного элемента над другим по отношению к свойству нечеткого множества. Парные сравнения удобно представлять следующей матрицей:

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix},$$

где a_{ij} – уровень преимущества элемента x_i над $x_j, i, j = 1, \dots, n$, определяемый, например, по девятибалльной шкале Саати:

- 1 – если отсутствует преимущество элемента x_i над элементом x_j ;
- 3 – если имеется слабое преимущество x_i над x_j ;
- 5 – если имеется существенное преимущество x_i над x_j ;
- 7 – если имеется явное преимущество x_i над x_j ;
- 9 – если имеется абсолютное преимущество x_i над x_j ;
- 2, 4, 6, 8 – промежуточные сравнительные оценки.

Матрица парных сравнений является диагональной ($a_{ii} = 1, i = 1, \dots, n$) и обратно симметричной ($a_{ij} = 1/a_{ji}, i, j = 1, \dots, n$). Степени принадлежности принимаются равными соответствующим координатам собственного вектора $W = (w_1, w_2, \dots, w_n)$ матрицы парных сравнений: $\mu(x_i) = w_i, i = 1, \dots, n$.

Собственный вектор находится из следующей системы уравнений:

$$\begin{cases} A \cdot W = \lambda_{max} \cdot W \\ w_1 + w_2 + \dots + w_n = 1, \end{cases} \quad (2)$$

где λ_{max} – максимальное собственное значение матрицы A . Если нечеткое множество получилось субнормальным, все степени принадлежности делятся на максимальное значение $\mu(x_i)_{max}$.

Отклонение λ_{max} от n может служить мерой несогласованности парных сравнений эксперта. При согласованных мнениях эксперта матрица парных сравнений обладает следующими свойствами:

- она диагональная, то есть $a_{ii} = 1, i = 1, \dots, n$;
- она обратно симметрична, то есть элементы, симметричные относительно главной диагонали, связаны зависимостью $a_{ij} = 1/a_{ji}, i, j = 1, \dots, n$;
- она транзитивна, то есть $a_{ik} a_{kj} = a_{ij}, i, j, k = 1, \dots, n$.

Наличие этих свойств позволяет определить все элементы матрицы парных сравнений, если известны $(n - 1)$ недиагональных элементов. Например, если известна k -я строка, то есть элементы $a_{kj}, i, j = 1, \dots, n$, то произвольный элемент a_{ij} определяется так:

$$a_{ij} = \frac{a_{kj}}{a_{ki}}, i, j, k = 1, \dots, n. \quad (4)$$

После определения всех элементов матрицы парных сравнений степени принадлежности нечеткого множества вычисляются по формуле:

$$\mu(x_i) = \frac{1}{a_{1i} + a_{2i} + \dots + a_{ni}}. \quad (5)$$

Формула (5) не требует выполнения трудоемких вычислительных процедур, связанных с нахождением собственного вектора матрицы A .

Выводы. На сегодняшний день методы обработки экспертной информации имеют большое практическое и научное значение. В работе проведен анализ проблемы обработки экспертной информации в задачах классификации объектов для количественных и качественных экспертных оценок. Перспективным является улучшение качества экспертных оценок путем учета влияния фактора несогласованности экспертов на процесс классификации и разработка шкалы оценок.

Литература

1. Статистические методы анализа экспертных оценок / [отв. ред. Т. В. Рябушкин и др.]. – М. : Наука, 1977. – Т. 29. – 385 с.
2. Литвак Б. Г. Экспертная информация. Методы получения и анализа / Литвак Б. Г. – М. : Радио и связь, 1982. – 184 с.
3. Ульяновська Ю. В. Аналіз основних аспектів побудови інтелектуальної автоматизованої системи ідентифікації творів мистецтва / Ю. В. Ульяновська // Вісник Академії митної служби України. – 2002. – № 1(13). – С. 70–74.
4. Золотухин О. А. Обработка данных в линейных шкалах / О. А. Золотухин, О. К. Илюнин // Автоматизированные системы управления и приборы автоматики : республиканский межведомственный научно-технический сборник. – 1987. – Вып. 83. – С 59–65.
5. Мороз Б. И. Анализ мер близости объектов для различных типов экспертной информации / Б. И. Мороз, Ю. В. Ульяновская // АСУ и приборы автоматики. – 2008. – Вып. 144. – С. 194–198.