УДК 007

А.Е. Архипов, С.А. Архипова

ИНФОРМАЦИОННЫЙ ПОДХОД К АНАЛИЗУ ДАННЫХ ЭКСПЕРТНОГО ОЦЕНИВАНИЯ

Введение

Одним из результатов процесса информатизации общества есть постоянный рост информационной нагрузки на субъектов информационной деятельности, в частности на лиц, которые по должности или в связи с другими обстоятельствами должны принимать решения (ЛПР – лица, принимающие решения). Учитывая динамизм и разнообразие вопросов, с которыми сталкиваются ЛПР, широкий профессиональный диапазон этих вопросов, который может значительно превышать границы компетенции ЛПР, все более распространенным становится привлечение для решения этих вопросов экспертов. Метод экспертного опроса, в частности групповая экспертиза, является оперативным и достаточно простым методом получения информации, но требует достаточно серьезного отношения к ее обработке.

В связи с этим актуальными становятся вопросы разработки общих подходов и методов анализа и обработки данных групповой экспертизы.

Постановка задачи

Рассматривается задача классификации группы объектов, составляющих некоторое множество $O = \{O_j\}, j = \overline{1, M}$, каждый из которых может быть отнесен к определенному классу (кластеру), являющемуся элементом множества $C = \{c_i\},\ i = \overline{1, n}$. Объекты классифицируются m экспертами, образующими множество $E = \{E_k\},\ k = \overline{1, m}$. Результаты классификации для произвольного объекта O_j представляются частотным распределением в виде вектора $X_j = [x_1, \ldots x_i, \ldots x_n]$, где x_i – количество экспертов, отнесших экспертируемый объект к классу c_i.

Требуется построить схему (алгоритм) обработки экспертных данных, позволяющую по результатам анализа частотных распределений $X_j,\ j = \overline{1, m}$ отнести соответствующие объекты к одному из классов (кластеров). Дополнительным условием, осложняющим решение задачи, является случайное количество экспертов, участвующих в экспертизе каждого из объектов.

Решение задачи

Случайный объем выборки экспертных оценок, полученных в ходе классификации произвольного объекта O_j, затрудняет применение наиболее часто используемых при решении подобных задач пороговых методов классификации [1]. Для решения поставленной задачи перспективным представляется использование аппарата теории информации.

На рис. 1 система получения и обработки экспертной информации представлена в виде многоканальной информационной системы (произвольный k-й эксперт представлен измерителем E_k), на вход которой подается общий для всех каналов сигнал, а вектор из m выходящих сигналов поступает в блок обработки БО, где реализуется обработка и обобщение результатов групповой экспертизы и формируется конечное решение С. Процедура обработки должна максимально учитывать особенности оценок, полученных от каждого отдельного эксперта, поэтому на первый план выходят вопросы выбора показателей и критериев, по которым формируется конечное решение С.

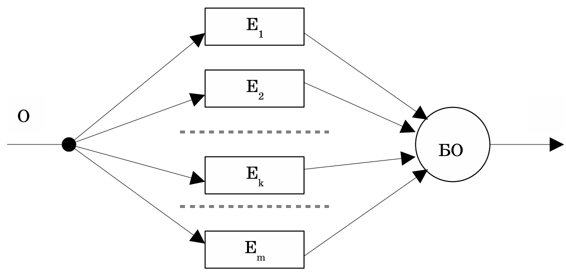

Рис. 1 – Схема получения и обработки экспертной информации

Введем понятие активного количества экспертов S_j:

$$S_j = \sum_{i=1}^{n} x_{ij}, \tag{1}$$

которое включает только тех экспертов, чьи оценки вошли в частотное распределение $X_j = [x_1, \dots x_i, \dots x_n]$, характеризующее объект O_j. Отметим, что возможны случаи, когда $m \neq S_j$. Например, в результате небрежности эксперта один объект оказывается отнесенным сразу к нескольким классам ($S_j > m$) либо наоборот, эксперт вообще “забывает” классифицировать объект ($S_j < m$), т.е. фактически выбывает из экспертизы.

Полагаем, что к моменту начала экспертизы какого-либо объекта вероятность отнесения его к любому классу $c_i, i = \overline{1, n}$ одинакова и равна $p_0 = 1/n$, а соответствующая априорная неопределенность классификации равна

$$H_0 = -\sum_{i=1}^{n} \frac{1}{n} \ln \frac{1}{n} = \ln n, \tag{2}$$

т.е. соответствует энтропии дискретной случайной величины C с n равновероятными исходами (реализациями) $c_1, \dots, c_n$.

 ISSN 1562-9945

После предъявления экспертам конкретного объекта O_j вероятности отнесения его к тому или иному классу определяются соответствующим элементом вектора X_j и в общем случае отличаются от исходного значения p_0: $p_{1j} = x_{1j}/S_j, \ldots, p_{nj} = x_{nj}/S_j$, т.е. в ходе экспертизы происходит уменьшение априорной энтропии H_0 до величины апостериорной энтропии

$$H_j = -\sum_{i=1}^{n} p_{ij} \ln p_{ij}. \tag{3}$$

Естественной количественной характеристикой степени снятия априорной неопределенности является разность энтропий

$$I_j = H_0 - H_j, \tag{4}$$

которую можно назвать экспертной информацией. Величина I_j, непосредственно определяющая степень объективности классификации объекта O_j, зависит от неравномерности апостериорного распределения экспертных оценок, входящих в вектор X_j.

В частности, если все эксперты одинаково классифицируют объект O_j , отнеся его к классу c_i, полученная экспертная информация целиком “снимает” исходную априорную неопределенность H_0, т.е. $H_j = 0, I_j = H_0 = I_{\max}$. Подобный исход является в определенной степени уникальным, идеализированным. Для реальной экспертизы обычно характерна многоальтернативность вариантов классификации, где приоритет отдается варианту c_j, имеющему наибольшую частоту выбора $x_i = x_{\max}$.

Однако информация I_j не является единственным и достаточным фактором, обеспечивающих возможность объективной классификации. Представима ситуация, при которой существенно уменьшается степень исходной априорной неопределенности, однако исключена возможность окончательной классификации объекта. Пример этого – экспертиза, по результатам которой исходная неопределенность H_0 сокращается до апостериорной $H_j = \ln 2$, соответствующей двум оставшимся альтернативным равновероятным вариантам классификации. Несмотря на то, что при больших n информация $I_j = \ln(n/2)$ может быть достаточно большой, появление подобных равновероятных альтернатив заводит экспертизу в тупик.

Поэтому помимо учета величины экспертной информации I_j для обоснования конечных выводов классификационной экспертизы необходимо изучить особенности распределения значений элементов вектора X_j.

Пусть данные экспертизы конкретного объекта O_j составляют последовательность $x_1, \ldots, x_q, \ldots, x_g, \ldots, x_n$, ранжирование которой приводит к ряду $x_{\max} = x_{(1)} \geqslant x_{(2)} \geqslant \ldots \geqslant x_{(n)} = x_{\min}$, причем соответствие индексов до и после ранжирования для двух первых элементов ранжированного ряда определяется соотношениями: $x_q = x_{(1)}, x_g = x_{(2)}$. Если $x_{(1)} > x_{(2)}$, классификация возможна, причем степень обоснованности ее результата (принадлежность объекта O_j классу C_q) можно оценить, сравнив произошедшие в ходе экспертизы изменения частных неопределенностей для классов c_qи c_g и сопоставив их с общим количеством экспертной информации I_j. Учитывая, что априорная частная неопределенность классификации для произвольного класса c_i равна $h_0 = -\ln p_0 = \ln n$, а апостериорные неопределенности для классов c_q, c_g соответственно составляют $h_{qj} = -\ln p_{qj},\ h_{gj} = -\ln p_{gj}$, где $p_{qj} = x_{qj}/S_j,\ p_{gj} = x_{gj}/S_j$ – относительные частоты выбора экспертами классов c_q, c_g, получим выражение для соответствующих частных информаций:

$$I_{qj} = h_0 - h_{qj} = \ln \frac{nx_{qj}}{S_j} = \ln \frac{p_{qj}}{p_0}, \tag{5}$$

$$I_{gj} = h_0 - h_{gj} = \ln \frac{nx_{gj}}{S_j} = \ln \frac{p_{gj}}{p_0}. \tag{6}$$

В [2] выражение вида $\ln(p/p_0)$, где p_0, p – вероятности достижения некоторой цели до и после поступления информации, называется мерой ценности информации. При наличии нескольких альтернатив (например, при классификации) решение принимается в пользу той, для которой мера ценности имеет большее численное значение. Рассмотрим усредненные частные информации вида $p\ln(p/p_0)$. Для них справедливо равенство

$$\sum_{i=1}^{n} p_{ij} J_{ij} = \sum_{i=1}^{n} p_{ij} \ln \frac{p_{ij}}{p_0} = H_0 - H_j = I_j. \tag{7}$$

Для оценки обоснованности и значимости выбираемого варианта классификации целесообразно ввести показатель качества классификации, базирующийся на сопоставлении нормированных частных информаций лучшего и последующего за ним (второго) варианта классификации (в данном случае это соответственно классы c_q и c_g :

$$R_j(q,g) = \frac{1}{I_j}(p_{qj} \ln \frac{p_{qj}}{p_0} - p_{gj} \ln \frac{p_{gj}}{p_0}) = p_{qj}\frac{J_{qj}}{I_j} - p_{gj}\frac{J_{gj}}{I_j} = r_{1j} - r_{2j} \tag{8}$$

Первый элемент в (8) позволяет сопоставить количество частной информации по первой (выбираемой) альтернативе с общим количеством экспертной информации, $r_{1j} > 0$. Второй элемент, в зависимости от соотношений значений x_{gj} и S_j/n, может быть как больше, так и меньше 0. В целом показатель $R_j(q,g)$ позволяет оценить информационную контрастность лучшей и второй по значимости альтернатив. Уровень этой контрастности определяет степень обоснованности принимаемого решения. Для идеального безальтернативного варианта классификации, когда $H_j = 0, I_j = I_{\max} = H_0, p_{qj} = 1, p_{gj} = 0$, коэффициент $R_j == r_{1j} = 1$. Зона малых значений $0.3 \leqslant R_j(q,g) < 0.4$ определяет зону неуверенной классификации, где окончательное решение должен принимать специалист по обработке. Варианты классификации, для которых $R < 0.3$, отбрасываются из-за недостаточной обоснованности первой альтернати-

 ISSN 1562-9945

вы по отношению к другим и в первую очередь ко второй альтернативе. В случае, когда $x_{qj} = x_{gj}$, значение $R_j(q, g) = 0$.

Выводы

Применение информационного подхода для описания процедуры групповой экспертизы приводит к достаточно простым математическим соотношениям, дающим представление о перераспределении неопределенности и информации в процессе экспертизы и позволяющим сформировать показатель качества классификации, эффективно применимый для окончательного принятия решения по имеющимся данным групповой экспертизы.

Литература

1. Прикладная статистика: Классификация и снижение размерности: Справ.изд./ С.А. Айвазян и др. – М.: Финансы и статистика. 1989. – 607с.

2. Харкевич А.А. О ценности информации. – В сб.: Проблемы кибернетики. – М.: Физматгиз, 1960, вып.4, с.53-57.