

МЕТОДИКА АВТОМАТИЗИРОВАННОГО СИНТЕЗА ФОРМАЛИЗОВАННЫХ ОПИСАНИЙ ЗНАНИЙ ДЛЯ РАСПОЗНАВАНИЯ ВОЗДУШНЫХ ОБЪЕКТОВ

к.т.н. Б.И. Низиенко, О.В. Шевченко, А.В. Александров
(представил д.т.н., проф. Е.И. Бобыр)

Рассматривается методика автоматизации процедур синтеза и корректировки структур знаний о процессах распознавания воздушных объектов и автоматизированного пополнения базы знаний для экспертных самообучающихся систем.

Постановка проблемы. Одной из наиболее трудоемких задач, возникающих при организации функционирования экспертных систем (ЭС) вообще и экспертных систем распознавания воздушных объектов в частности, является задача формализации знаний о процессах распознавания, пополнения и корректировки базы знаний. При этом наиболее предпочтительным для использования в автоматизированных системах военного назначения представляются ЭС, в которых предусмотрена частичная или полная автоматизация процедур корректировки структур знаний о процессах распознавания воздушных объектов – самообучающиеся экспертные системы (СЭС). Для таких систем процесс синтеза формализованных описаний знаний о процессах распознавания воздушных объектов можно разбить на ряд этапов:

- 1) определение экспертами характера (вида) признаков, используемых в процессе распознавания (количественных или качественных);
- 2) выбор списка рабочих алфавитов воздушных объектов, в которых будет осуществляться распознавание, а также классов самих распознаваемых объектов в каждом из алфавитов;
- 3) автоматизированное формирование СЭС на основе обучающей выборки априорных распределений значений признаков для каждого из объектов, относящихся к распознаваемым классам;
- 4) выявление степени информативности признаков и, на основе этого, структуризация формализованных описаний классов распознаваемых объектов.

В статье предлагается подход к решению последней задачи.

Анализ литературы. В целом ряде работ в качестве формально-логического аппарата представления знаний об описании классов распознаваемых объектов предложено использовать комбинированную модель

знаний, построенную на основе теории возможностей, нечетких множеств и многозначной логики [1, 2].

В [3] предложено использование нечетких множеств для формализации структур знаний о закономерностях проявления количественных признаков с нестохастической неопределенностью. Знания о значениях качественных признаков для описания классов распознаваемых объектов предложено формализовать в виде распределений возможностей значений признаков для каждого из классов алфавита. В обобщенном виде формализация знаний о значениях признаков производится в соответствии с выражением

$$X = \begin{cases} \mu(x), & \text{аннèè ð - èîèè÷-âñðââá íúé òðèçíâè ;} \\ \pi(\hat{E}_i), & \text{аннèè ð - èâ÷-âñðââááí úé òðèçíâè ,} \end{cases} \quad (1)$$

где $\mu(x)$ – функция принадлежности значений количественного признака нечеткому множеству описания закономерностей его проявления для соответствующего класса; $\pi(K_i)$ – распределение возможностей значений качественно признака для соответствующих классов алфавита.

При формализации количественных признаков вид функции принадлежности может быть произвольным. В [4] предложено использовать комбинированную модель на основе теории возможностей и нечетких множеств для категорирования воздушных объектов с использованием нечетких L-R интервалов. Однако использование такого подхода не всегда позволяет адекватно представить описания соответствующих классов воздушных объектов. Вместе с тем, к формальному представлению признаков необходимо предъявить ряд требований: простота построения формального описания с точки зрения пользователя; высокая выразительность с точки зрения описательных возможностей; удобство последующей обработки на уровне программных модулей; удовлетворительная робастность.

С учетом этих требований целесообразным является использование для формализации знаний о закономерностях проявления значений количественных признаков распознаваемых объектов нечетких множеств, представленных в виде гистограмм. Такой подход позволяет с удовлетворительной степенью достоверности представить соответствующие знания и обеспечивает простоту автоматизации процедур их обработки.

Цель статьи. Представление методики автоматизированного синтеза формализованных описаний знаний для распознавания воздушных объектов, использующей для формализации знаний о закономерностях проявления значений признаков распознаваемых объектов нечетких множеств и интервалов с использованием многозначных дискретных значений.

Основная часть. При построении формализованного описания классов должна быть учтена логическая взаимосвязь признаков. Так как

любое выражение на языке исчисления предикатов можно представить в стандартном виде, то и описание класса распознаваемого объекта на языке булевой алгебры можно представить в виде:

$$K = \bigvee_{i \in I} \bigwedge_{j \in J} X_{ij}, \quad (2)$$

где X_{ij} – переменные, описывающие некоторые понятия, утверждения [3].

В качестве таких понятий предлагается использовать утверждения (либо их отрицания) о подтверждении полученной оценки значения признака X_{ij} его априорному описанию в рассматриваемом классе.

Для последующих рассуждений будем считать, что в результате выполнения первых трех этапов обозначенной выше процедуры синтеза формализованного описания классов воздушных объектов сформированы законы распределения значений признаков для каждого класса алфавитов.

Для определения степени информативности признаков распознаваемых объектов необходимо провести оценку их прагматической ценности (степени «полезности» с точки зрения использования для процесса распознавания).

Предложено все признаки разделить на две группы – существенные и несущественные признаки распознаваемых объектов. На содержательном уровне этот процесс представляет собой выявление признаков, содержащих наиболее ценные знания о распознаваемых объектах и признаках, которые лишь косвенно могут свидетельствовать в пользу принадлежности наблюдаемого объекта к распознаваемому классу. При рассмотрении частного случая распознавания объекта в алфавите из двух классов с прагматической точки зрения *признак будет тем информативнее, чем больше будут различаться между собой его описания для каждого из классов алфавита. При произвольном количестве классов исходного алфавита степень информативности признака в описании класса следует рассматривать как наибольшую степень отличия при парных сравнениях с описаниями этого признака во всех других классах алфавита.*

В общем случае в ЭС распознавания воздушных объектов некорректно говорить об однозначности принимаемых решений [5]. Поэтому для определения степени информативности признаков распознаваемых объектов воспользуемся модальностями. Такие модальности должны позволить выразить неоднозначное и, в то же время, близкое отношение рассматриваемых понятий (с точки зрения их соответствия физической реальности) тем, которые используются в ходе мыслительной деятельности человека при решении рассматриваемых задач.

Существенными для распознавания принадлежности объекта некоторому классу будем считать признаки, обнаружение которых в значительной степени свидетельствует о принадлежности данного воз-

душного объекта к рассматриваемому классу соответствующего алфавита. Если степень информативности признака для решения задачи распознавания высока, то его следует отнести к разряду существенных.

Несущественными для распознавания принадлежности объекта некоторому классу будем считать признаки, обнаружение которых свидетельствует о возможной (косвенной) принадлежности объекта одновременно к нескольким классам соответствующего алфавита. Несущественными следует считать малоинформативные признаки.

Для определения степени информативности признаков для распознавания объектов введем ее количественную оценку. В качестве такой оценки может быть использована мера близости (толерантности) двух распределений значений признака в описаниях классов одного алфавита.

Пусть X_i и X_j – описания закономерностей проявления признака X для объектов классов K_i и K_j соответственно. Пусть $An(X_i, X_j)$ – мера близости двух распределений X_i и X_j , удовлетворяющая свойствам:

$$S_1 : An(X_i, X_j) \in [0,1] ; \quad (3)$$

$$S_2 : An(X_i, X_j) = 1, \text{ если распределения } X_i \text{ и } X_j \text{ идентичны ;} \quad (4)$$

$$S_3 : An(X_i, X_j) = 0, \text{ если два распределения не пересекаются .} \quad (5)$$

Введем формально правила определения степени информативности признаков, а также правила селекции существенных и несущественных признаков в описаниях классов некоторого алфавита. Ввиду разнородности признаковой информации и математических методов формализации их априорных описаний, эти правила для количественных и качественных признаков будут отличаться.

При рассмотрении количественных признаков сравнению подлежат два закона распределения возможностей. В теории статистических решений существуют методы определения степени толерантности двух распределений. Эти методы, в основном, используют различные критерии согласия. С учетом особенностей формализации закономерностей проявления признаков, в качестве такого критерия согласия предлагается использовать меру, удовлетворяющую свойствам $S_1 \dots S_3$:

$$An(X_i, X_j) = 1 - \frac{\sum_{n=1}^N |\pi_1(n) - \pi_2(n)|}{N} , \quad (6)$$

где $\pi_1(n)$ и $\pi_2(n)$ – меры возможности значений признака X , попадающих в элементарный интервал с номером n ; N – суммарное количество элементар-

ных интервалов, попадающих в отрезок $[n_{\min}^1, n_{\max}^2]$ описаний законов распределений значений признака X для классов K_1 и K_2 соответственно.

В соответствии с принятым методом формализации качественные признаки представлены в виде распределений возможностей для классов алфавитов $\pi(K_i)$. Для распределения возможностей принадлежности объекта некоторым классам по значению качественного признака может быть построено соответствующее распределение базисных вероятностей $m(E_k)$ фокальных базовых элементов E_k для множества классов [2].

Для двух распределений базисных вероятностей существует показатель степени их несогласованности m_0 :

$$m_0 = \sum_{E_i \cap E_j = 0} m_1(E_i) \cdot m_2(E_j) . \quad (7)$$

Значение m_0 отражает степень противоречивости двух распределений базисных вероятностей друг относительно друга. Иллюстрация процесса определения степени близости двух априорных распределений значений признака представлена на рис. 1.

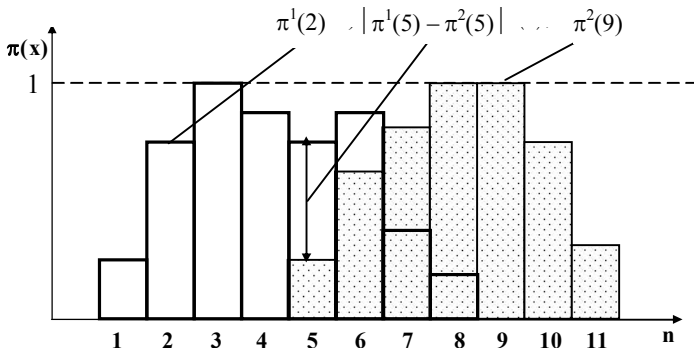


Рис. 1. Иллюстрация к порядку определения степени близости двух распределений значений количественного признака

В качестве степени близости качественных признаков предлагается принять величину

$$A_n(X_i, X_j) = 1 - m_0, \quad (8)$$

где m_0 – степень несогласованности распределений базисных вероятностей.

Если одно распределение возможностей классов качественного признака таково, что соответствующее ему распределение базисных вероятностей $m_1(E_k)$ полностью согласовано ($m_0 = 0$) с распределением базисных вероятностей, соответствующим другому описанию этого же признака $m_2(E_i)$, то можно считать эти описания идентичными и степень их близости $A_n(X_1, X_2) =$

= 1. В случае полной несогласованности ($m_0 = 1$), рассматриваемые описания не подтверждают друг друга и не совпадают. С аксиоматической точки зрения использование степени несогласованности является корректным, так как при этом выполняются равенства (3...5).

Основываясь на изложенных определениях существенных и несущественных признаков в описаниях классов алфавита, формально правила их селекции можно определить следующим образом:

$$\forall_{i=1..M, i \neq j}, A_n(X_i, X_j) \leq A_{n_п\ddot{o}} \rightarrow X_j \text{ (существенный признак в описании класса } K_j \text{)}; \quad (9)$$

$$\forall_{i=1..M, i \neq j}, A_n(X_i, X_j) > A_{n_п\ddot{o}} \rightarrow X_j \text{ (несущественный признак в описании класса } K_j \text{)}. \quad (10)$$

При этом пороговые значения степеней близости $A_{n_пор}$ выбираются экспертным либо экспериментальным путем в период отладки процедур синтеза структуры базы знаний СЭС. С учетом минимума среднего риска ошибок оценивания первичное значение порога может быть выбрано равным 0,5.

На этапе синтеза формализованного описания знаний о классах распознаваемых объектов полученные закономерности проявления признаков должны быть структурированы в соответствии со значением степени информативности признаков и выбранным формально-логическим аппаратом представления знаний. Формирование формализованного описания знаний о процессах распознавания воздушных объектов заключается в определении вида логических отношений между признаками (группами признаков).

Процедура автоматизации процесса структуризации формализованных описаний признаков может быть основана на следующих принципах.

1. Решение о классе распознаваемого объекта может быть с необходимой степенью достоверности принято при обнаружении хотя бы одного наблюдаемого существенного признака.

2. При отсутствии существенных признаков для распознавания необходимо использовать все множество наблюдаемых несущественных признаков.

В соответствии с 1-м принципом, все существенные признаки (группы признаков) в описании класса K_m целесообразно группировать операцией объединения

$$\forall i, \text{Тип}(X_i) = \{\text{существенный}\} \rightarrow K_m = \bigvee_{i \in S} X_i, \quad (11)$$

где S – множество существенных признаков описания объектов.

В соответствии со 2-м принципом все несущественные признаки (группы признаков) целесообразно группировать операцией пересечения

$$\forall j, \text{Тип}(X_j) = \{\text{несущественный}\} \rightarrow Z_m = \bigwedge_{j \in N} X_j, \quad (12)$$

где N – множество несущественных признаков рассматриваемого класса.

В целом, описание класса посредством выявленных закономерностей проявления признаков может быть представлено в виде

$$K_m = \left(\bigvee_{\forall i \in S} X_i \right) \bigvee \left(\bigwedge_{\forall j \in N} X_j \right). \quad (13)$$

Роль эксперта при таком подходе сводится к корректировке синтезированных формализованных описаний и к выбору пороговых значений для решающего правила селекции существенных и несущественных признаков.

Выводы. Рассмотренная методика позволяет автоматизировать процесс синтеза формализованных описаний классов распознаваемых объектов. Приведенная форма представления знаний позволяет автоматизировать процесс выявления самих закономерностей проявления признаков для каждого из объектов распознавания и минимизировать время привлечения экспертов к процессу синтеза соответствующих структур знаний и пополнения баз знаний СЭС.

ЛИТЕРАТУРА

1. Дюбуа Д., Прад А. *Теория возможностей. Приложения к представлению знаний в информатике: пер. с фр.* – М.: Радио и связь, 1990. – 212 с.
2. Грачев В.М., Попрыгин А.Н. *Методика распознавания классов воздушных объектов в АСУ ПВО с использованием однородной функциональной сети // Сб. научн. тр. ХВУ. – Х.: ХВУ. – 1995. – Вып. 8. – С. 49 – 54.*
3. Щербаков В.И., Низиенко Б.И., Попрыгин А.Н., Шевченко О.В. *Использование новых информационных технологий для диагностики заболеваний в неотложной хирургии. – Х.: Бизнес информ, 2000. – 202 с.*
4. Низиенко Б.И., Затхей В.А., Селезнев С.Е. *Использование функциональных сетей для означивания начальных условий в открытых экспертных системах // Зб. наук. пр. ХВУ. – Х.: ХВУ. – 2000. – Вып. 4(30). – С. 126 – 131.*
5. Ярушек В.Е. *Теоретические основы автоматизации процессов выработки решений в системах управления. – Х.: ХВУ, 1993. – 446 с.*

Поступила 9.10.2003

НИЗИЕНКО Борис Иванович, канд. техн. наук, доцент, нач. каф. ХВУ. В 1981 году окончил ХАИ, в 1992 – ВИРТА ПВО. Область научных интересов – применение методов искусственного интеллекта для автоматизации процессов управления войсками и оружием.

ШЕВЧЕНКО Олег Владимирович, начальник НИЛ кафедры ХВУ. Область научных интересов – применение методов искусственного интеллекта для автоматизации процессов управления войсками и оружием.

АЛЕКСАНДРОВ Александр Валерьевич, адъюнкт ХВУ. Окончил ХВУ в 1999 году. Область научных интересов – применение методов искусственного интеллекта для автоматизации процессов управления войсками и оружием.