Представление прецедентов и их поиск в базе знаний на основе теории грубых множеств

Авторы: *Коваленко И.И., Мельник А. В., Национальный университет кораблестроения имени адмирала Макарова*

Важным структурным элементом системы искусственного интеллекта являются базы знаний (БЗ), которые формируются в процессе представления и структурирования данных о различных предметных областях. Традиционными методологиями структурирования знаний являются объектно-ориентированный и объектно-структурный подходы, которые, наряду с общеизвестными положительными характеристиками, обладают существенным недостатком. Такие подходы для представления знаний используют строгие модели, определённые релевантностью (определённостью) рассматриваемых элементов, что на практике заставляет разработчиков урезать реальные знания экспертов.

В настоящее время сформировался и интенсивно развивается альтернативный подход для формирования БЗ – метод рассуждений на основе прецедентов (CBR – Case-Based Reasoning), который представляет собой процесс получения решений новых проблем на основе решения аналогичных проблем в прошлом [1].

В формализованном виде прецедент (case) представляется в виде набора параметров с конкретными значениями:

$$ПРЕЦЕДЕНТ\ (case)(x_1, x_2, ..., x_n, R), \quad (1)$$

где $x_1, ..., x_n$ – параметры ситуаций, описывающий данный прецедент $(x_1 \in X_1, x_2 \in X_2, ..., x_n \in X_n)$; R – рекомендации лицу, принимающему решение (ЛПР); $X_1, ..., X_n$ – область допустимых значений соответствующих параметров прецедента; n – количество параметров прецедента.

Основными задачами, которые решаются при формировании базы прецедентов (БП), является создание инструментальных средств их компонентного представления и поиска. Существующие решения указанных задач (например, представление прецедентов фреймовыми моделями или их поиск с применением метода ближайшего соседа) также не лишены недостатков, которые отображены в ряде публикаций [1, 2].

Вместе с тем, в последние годы получила развитие теория грубых множеств (ТГМ) [3], которая может внести свой вклад в решение рассмотренных выше проблем, связанных с формированием БЗ и БП. ТГМ рассматривается как концепция рассуждения о знаниях, когда они неточны (неточные знания).

Данная теория основана на том, что знания глубоко внедрены в способность людей выполнять классификацию предметов, явлений, объектов, ситуаций и др. Другими словами, знания основаны на способности классифицировать объекты. Рассмотрим одну из ситуаций, которая предопределяет необходимость ТГМ. Пусть, например, исходное множество $\{CASE\}$ представлено двумя классами $CASE_{S1} = \{x_1, x_2, x_3, x_4\}$ и $CASE_{S2} = \{x_5, x_6, x_7\}$, $CASE_{S1}, CASE_{S2} \subseteq CASE$. Пусть также сформирован новый прецедент $CASE_0 = \{x_3, x_4, x_7, x_8\}$, который необходимо отнести к одному из указанных классов S_1 и S_2. Однако, можно видеть, что элементы $\{x_3, x_4\} \in S_1$, $\{x_7\} \in S_2$, а элемент $\{x_8\} \notin (S_1, S_2)$. Рассмотренный пример характеризует наличие неопределённости, что влечёт за собой «неточную» классификацию. Тем не менее, такая ситуация на практике может выглядеть более реально, чем точная классификация.

В ТГМ БЗ определяется как $K = (U, R)$, где U – универсум элементов; R – отношение эквивалентности, на основе которого могут быть выделены классы эквивалентности (категории) элементов U, которые обозначаются через $IND(R)$. В каждую категорию включаются элементы, которые обладают одинаковыми значениями классификационных признаков (атрибутов). Внутри каждой категории элементы считаются неразличимыми. Пусть имеется исходный универсум элементов $U = \{x_1, x_2, ..., x_n\}$, тогда относительно классификации, $IND(R)$ могут существовать следующие ситуации [4]:

1. Множество X является объединением некоторых категорий из $IND(R)$, тогда оно называется R-точным.

2. Множество X не может быть выражено как объединение некоторых категорий $IND(R)$, тогда оно называется R-неточным или R-грубыми. В этом случае вводятся понятия R-нижней аппроксимации: $\underline{R}\mathrm{X} = \cup\{\mathrm{Y} \in IND(R) : \mathrm{Y} \subseteq \mathrm{X}\}$ и R-верхней аппроксимации. $\overline{R}\mathrm{X} = \cup\{\mathrm{Y} \in IND(R) : \mathrm{Y} \cap \mathrm{X} \neq \varnothing\}$.

R-нижнюю аппроксимацию множества X называют R-положительной областью: $POS_R(\mathrm{X}) = \underline{R}\mathrm{X}$. Отрицательной областью X называется подмножество элементов универсума, которые с определённостью не принадлежат X: $NEG_R(\mathrm{X}) = U - \overline{R}\mathrm{X}$.

Граничной областью X называют подмножество его элементов, которые принадлежат R-верхней аппроксимации, но не принадлежат R-нижней аппроксимации: $BN_R(\mathrm{X}) = \overline{R}\mathrm{X} - \underline{R}\mathrm{X}$.

Рассмотрим пример, иллюстрирующий рассмотренные основные понятия ТГМ. Имеется база знаний $K = (U, R)$, $U = \{x_1, ..., x_{10}\}$ – универсум элементов, R – отношение эквивалентности, на основе которого выделены следующие классы эквивалентности (категории) не U: $U / IND(R) = \{\{x_1, x_2\}, \{x_3, x_7, x_{10}\}, \{x_4\}, \{x_5\}, \{x_6\}, \{x_8\}, \{x_9\}\}$. Заданы целевые множества $\mathrm{X}_1 = \{x_1, x_2, x_4, x_5\}$, $\mathrm{X}_2 = \{x_1, x_2, x_3, x_4\}$, для которых необходимо определить аппроксимации, отрицательные и граничные области. Имеем: $\underline{R}\mathrm{X}_1 = \{x_1, x_2, x_4, x_5\}$, $\overline{R}\mathrm{X}_1 = \varnothing$, $NEG_R(\mathrm{X}_1) = \{x_3, x_6, x_7, x_8, x_9, x_{10}\}$, $BN_R(\mathrm{X}_1) = \varnothing$; $\underline{R}\mathrm{X}_2 = \{x_1, x_2, x_4\}$, $\overline{R}\mathrm{X}_2 = \{x_1, x_2, x_3, x_4, x_7, x_{10}\}$, $NEG_R(\mathrm{X}_2) = \{x_5, x_6, x_8, x_9\}$, $BN_R(\mathrm{X}_2) = \{x_3, x_7, x_{10}\}$.

В заключение необходимо отметить, что философия грубых множеств такова, что выделение релевантных категорий элементов на универсуме, характеризация целевых множеств и операции над этими множествами производятся только и только на основе существующих знаний [4]. Эти знания включают в себя значения атрибутов и отношения эквивалентности, на основе которых выделяются категории элементов.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Еременко Т. К. Использование CBR-подхода для баз знаний ситуационных центров / Т. К. Еременко, Ю. Г Пилипенко // Матер. конф. «Системи підтримки прийняття рішень. Теорія і практика», 2010. – С. 151–153.
2. Шерстюк В. Г. Формальная модель гибридной сценарно-прецедент-ной СППР / В. Ф. Шерстюк // Информационно-управляющие комплексы и системы, 2004. – №1 (13). – С. 1–11.
3. Pawlak Z. Rough Sets Theoretical Aspects of Reasoning about Data / Z. Pawlak. – Boston; London: Academic Publishers, 1991. – 229 p.

4. Uzga-Rebrovs O. Nenoteiktibu parvaldisana / O. Uzga-Rebrovs. – Rezekne: RA Izdevnieciba. 2010. – vol. 3. – 560 lpp.