

Универсальный метод учета ограничений существования свойств в задачах вывода формальных понятий из нечетких объектно-признаковых данных

А.Е. Самойлов^{1,2}, С.В. Смирнов¹

¹Самарский федеральный исследовательский центр РАН, Институт проблем управления сложными системами РАН, Садовая 61, Самара, Россия, 443020

²Самарский национальный исследовательский университет им. академика С.П. Королева, Московское шоссе 34а, Самара, Россия, 443086

Аннотация

Объектом исследований является процесс обработки эмпирически установленного соответствия «объекты-признаки (свойства)» - с целью выявления множества бикластеров, именуемых формальными понятиями, и построения решетки, фиксирующей частично упорядоченное отношение обобщения на множестве формальных понятий. Статья посвящена реализации предложенного ранее авторами метода предобработки исходного нечеткого соответствия «объекты-свойства», который обеспечивает в задаче вывода формальных понятий учет так называемых «ограничений существования свойств». Метод пригоден для всех типов задач бикластеризации: вывода четких и нечетких понятий из порогового сечения исходных данных, вывода нечетких понятий на основе нечеткого замыкания исходных данных. Реализация метода предусматривает представление нечеткого множества признаков каждого объекта обучающей выборки в виде композиции конечного набора четких множеств, фильтрацию этого набора множеств для учета ограничений существования свойств и аппроксимацию нечеткого множества признаков объекта обучающей выборки на основе композиции отфильтрованных множеств признаков. Эффективность предложенной реализации обеспечивается рациональной организацией каждого из указанных действий и фактически их частичным совмещением.

Ключевые слова

Анализ формальных понятий, ограничения существования свойств, нечеткий формальный контекст, нормальное множество свойств

1. Введение

К формированию знаний о предметной области (ПрО) в виде понятий сосуществуют два математически обоснованных подхода. В анализе формальных понятий (АФП) [1] бикластеры-понятия извлекаются из эмпирически установленного для ПрО соответствия «объекты-свойства», а авторы [2] предлагают находить содержания формальных понятий как нормальные подмножества набора свойств, характеризующих объекты ПрО, в результате анализа ограничений существования свойств (ОСС) - бинарных отношений обусловленности и несовместимости свойств. В [3] показана необходимость совмещать эти методы.

В [4] проанализирована проблема учета ОСС в АФП при различной постановке задачи вывода формальных понятий из нечетких объектно-признаковых данных: вывода нечетких понятий с помощью порогового сечения исходных данных, вывода нечетких понятий на основе нечеткого замыкания исходных данных. Предложен новый универсальный метод, основанный на теореме декомпозиции нечетких множеств. Здесь представлена эффективная реализация этого метода.

2. Эффективный алгоритм учета ограничений существования свойств

Пусть G - обучающая выборка объектов, M - множество измеряемых у объектов свойств, $I: G \times M \rightarrow [0, 1]$ – нечеткое соответствие «объекты-свойства», а $C: M \times M \rightarrow \{\mathbf{True}, \mathbf{False}\}$ и $E: M \times M \rightarrow \{\mathbf{True}, \mathbf{False}\}$ обусловленность и несовместимость свойств. Тогда:

1. Для каждого объекта $g \in G$ вычисляется значение r по формуле

$$r = \max_{m_k, m_j: E(m_k, m_j)} \min\{I(g, m_k), I(g, m_j)\},$$

затем вычисляется множество A по формуле

$$A = (0, r] \cup \bigcup_{m_k, m_j: C(m_k, m_j) \wedge I(g, m_k) > I(g, m_j)} (I(g, m_k), I(g, m_j)].$$

2. Для каждого свойства $m \in M$ проверяется, попадает ли значение $I(g, m)$ в один из интервалов множества A .
 - а) Если не попадает, то в соответствующую клетку нормализованного контекста вставляется число $I(g, m)$.
 - б) Если попадает в интервал $(a, b]$, то в соответствующую клетку нормализованного контекста вставляется число a .

3. Заключение

Эффективность предложенной реализации универсального метода учета ограничений существования свойств в задачах вывода формальных понятий из нечетких объектно-признаковых данных обеспечивается рациональной организацией каждого акта трансформации данных и фактически частичным совмещением некоторых актов.

Научная значимость полученного результата состоит в повышении адекватности нечеткого анализа формальных понятий, практическая - в обеспечении возможности решать задачи анализа данных большой размерности.

4. Литература

- [1] Ferré, S. Formal Concept Analysis: From Knowledge Discovery to Knowledge Processing / S. Ferré, M. Huchard, M. Kaytoue, S.O. Kuznetsov, A. Napoli // A Guided Tour of Artificial Intelligence Research. Vol. II: AI Algorithms. – Springer International Publishing, 2020 – P. 411-445.
- [2] Lammari, N. Building and maintaining ontologies: a set of algorithms / N. Lammari, E. Metais // Data & Knowledge Engineering. – 2004. – Vol. 48(2). – P. 155-176.
- [3] Смирнов, С.В. Две методологии вывода формальных понятий: когда и как они должны работать вместе / С.В. Смирнов // Знания – Онтологии – Теории: Материалы VII международной конференции. – Новосибирск: Институт математики СО РАН, Новосибирский государственный ун-т, 2019. – С. 355-363.
- [4] Самойлов, А.Е. Ограничения существования свойств в нечетком анализе формальных понятий / А.Е. Самойлов, С.В. Смирнов // Сборник трудов VI Международной конф. и молодежной школы «Информационные технологии и нанотехнологии» (ИТНТ). – Самара, 2020. – Т. 4. – С. 779-785.