Ограничения существования свойств в нечетком анализе формальных понятий

А.Е. Самойлов[1,2], С.В. Смирнов[2]

[1]Самарский национальный исследовательский университет им. академика С.П. Королева, Московское шоссе 34А, Самара, Россия, 443086
[2]Институт проблем управления сложными системами РАН - обособленное подразделение Самарского федерального исследовательского центра РАН, Садовая 61, Самара, Россия, 443020

Аннотация. Областью исследований является анализ нечетких объектно-признаковых данных с целью нечеткой бикластеризации, т.е. построение решеток нечетких формальных понятий. В центре внимания находится вопрос учета при выполнении анализа экзистенциональных отношений на множестве наблюдаемых и/или измеряемых признаков. Эти отношения известны субъекту-аналитику априори как «ограничения существования свойств» объектов исследуемой предметной области. Показано, что для наиболее популярного и практичного из двух основных методов построения решеток нечетких формальных понятий – «метода одностороннего порога доверия» - потенциальные нарушения ограничений существования свойств в решетке парируются методом рационального порогового сечения исходного нечеткого соответствия «объекты-свойства», ранее разработанным для вывода четких формальных понятий из нечетких исходных данных. Для второго метода построения нечетких формальных понятий – «нечеткого оператора замыкания» - подобный способ в принципе неприемлем. Для этого случая выдвигается идея специальной предварительной обработки исходных данных - «нормализации» нечеткого набора свойств для каждого объекта в обучающей выборке. Практическая ценность полученных результатов состоит в повышении адекватности приложений нечеткого анализа формальных понятий.

1. Введение

Одним из мощных методов интеллектуального анализа данных последних двух десятилетий стал анализ формальных понятий (АФП) [1-3]. Это прикладная ветвь алгебраической теории решеток, которая отражает классическое представление о понятии [4, 5] - фундаментальном элементе мышления, определяемом объемом (т.е. объектами, к которым применяется понятие) и содержанием (т.е. свойствами, охватываемые понятием). В АФП объемы и содержания сопряжены отношением I между множеством объектов G и множеством свойств M, I: $G\times M \to \{\mathbf{True}, \mathbf{False}\}$. Кортеж (G, M, I), исходно представляемый для анализа в форме таблицы соответствия «объекты-свойства», именуется формальным контекстом (ФК). ФК индуцирует операторы Галуа «$\uparrow$» и «$\downarrow$», а формальное понятие определяется как бикластер (X, Y), образуемый $X \subseteq G$ (объем) и $Y \subseteq M$ (содержание), удовлетворяющий $X\uparrow = Y$ и $Y\downarrow = X$, где $X\uparrow = \{m \in M \mid \forall g \in X: I(g, m) = \mathbf{True}\}$ и $Y\downarrow = \{g \in G \mid \forall m \in Y: I(g, m) = \mathbf{True}\}$. Множество всех формальных понятий, извлекаемых из ФК, упорядочено по включению объемов (или, равно, содержаний) и образует полную решетку, называемую решеткой понятий.

Нечеткий АФП (НАФП) – это, с одной стороны, адаптация АФП к доказанной психологами эластичности «человеческих понятий» в том смысле, что вопрос о том, применимо ли понятие к данному объекту, является для людей вопросом степени, а не вопросом «да»/«нет», и люди продуктивно работают именно в условиях такой грануляции мнений [6, 7]. С другой стороны, к НАФП приводит учет реалий накопления эмпирической информации, когда оценка истинности суждений вида «объекту $g \in G$ присуще свойство $m \in M$» расплывчата - например, является приближенной экспертной оценкой или формируется путем совмещения сведений из конкурирующих и противоречащих друг другу источников и т.п., - и для ее указания истинность нуждается в грануляции [8]. Формально и то и другое сводится к тому, что I: $G \times M \to L$, где L – многозначная шкала истинности. Собственно именование «нечеткий АФП» связано с выбором шкалы $L = [0, 1]$, используемой в нечеткой логике Л.А. Заде.

Еще одно развитие АФП связано с осмыслением его гипотетико-дедуктивной природы. В отличие от апостериори фиксируемого множества G набор измеряемых свойств объектов M формируется субъектом анализа априори и, как показано в [9], фактически является продуктом выдвижения гипотез о понятийном описании исследуемой предметной области. При этом определяется не только состав множества M, но и экзистенциональные отношения на этом множестве, известные как ограничения существования свойств (ОСС) [10-12]. Для вывода четких формальных понятий из нечеткого ФК при наличии ОСС в [12] предложен метод рационального α-сечения соответствия I.

Цель этой статьи – показать способы учета ОСС при выводе нечетких формальных понятий.

2. Ограничения существования свойств

Согласно [9] субъект АФП априори формирует систему измеряемых свойств, т.е. устанавливает состав множества M и два (и только два) экзистенциональных отношения на этом множестве:

- обусловленность C: $M \times M \to \{\mathbf{True}, \mathbf{False}\}$, когда наперед устанавливается, что обладая свойством m_j, всякий объект g непреложно обладает свойством m_k (хотя обратное может быть неверно), т.е. $C(m_j, m_k) \leftrightarrow \forall g \in G$: $m_j \in \{g\}\uparrow \to m_k \in \{g\}\uparrow$. Обусловленность рефлексивна, несимметрична и транзитивна;
- несовместимость E: $M \times M \to \{\mathbf{True}, \mathbf{False}\}$, когда, предопределяется, что обладая свойством m_j, всякий объект g заведомо не обладает свойством m_k, и наоборот, т.е. $E(m_j, m_k) \leftrightarrow \forall g \in G$: $m_j \in \{g\}\uparrow \to m_k \notin \{g\}\uparrow$. Отношение E антирефлексивно, симметрично и нетранзитивно, но характеризуется так называемой транзитивностью относительно обусловленности, что означает $\forall\ a, b, c \in M$: $C(a, b)\ \&\ E(b, c) \to E(a, c)$.

Эти отношения ограничивают состав или (со)существование свойств у объектов обучающей выборки: $g \in G$ может обладать лишь «нормальным» подмножеством множества измеряемых свойств M [11]. Подмножество измеряемых свойств $Y \subseteq M$ нормально тогда и только тогда, когда оно замкнуто и совместимо:

- Y замкнуто, если оно содержит все свойства, обусловленные любым элементом Y, т.е. $\forall m_j \in Y$: $(\exists m_k \in M$: $C(m_j, m_k)) \to m_k \in Y$;
- Y совместимо, если любые два элемента Y не связаны отношением несовместимости, т.е. $\forall m_j \in Y$: $(\exists m_k \in M$: $E(m_j, m_k)) \to m_k \notin Y$.

3. Два основных метода построения решеток нечетких формальных понятий

На сегодня в НАФП разработаны несколько методов построения решеток нечетких формальных понятий, а также их различные модификации (см., например, обзор [13], статьи [14-19]). К наиболее апробированным относят кратко очерчиваемые далее метод одностороннего порога [20-22] и метод, использующий оператор нечеткого замыкания [23].

3.1. Метод одностороннего порога

«Односторонность» метода состоит в том, что нечеткий ФК интерпретируется асимметрично как совокупность нечетких множеств над универсумом M, каждое из которых описывает один из объектов ФК. Иначе говоря, каждый $g \in G$ представляется в ФК нечетким множеством $\{I(g, m_1)/m_1,\ I(g, m_2)/m_2, \ldots, I(g, m_n)/m_n\}$, где $n = |M|$, $I(g, m_j)$ – степень истинности суждения «объекту g присуще свойство $m_j \in M$». Обычное (четкое) формальное понятие определяется в таком случае с помощью порогового значения $\alpha \in [0, 1]$ как пара (X, Y), $X \subseteq G$, $Y \subseteq M$, удовлетворяющая условиям $X\!\uparrow = Y$ и $Y\!\downarrow = X$, где $X\!\uparrow = \{m \in M \mid \forall g \in X{:}\ I(g, m) \geq \alpha\}$ и $Y\!\downarrow = \{g \in G \mid \forall m \in Y{:}\ I(g, m) \geq \alpha\}$.

Конструктивно на первом шаге метода производится α-аппроксимация нечеткого соответствия I, и далее рассматривается четкий ФК $(G, M, I^{(\alpha)})$, где

$$I^{(\alpha)}(g,m) = \begin{cases} \mathbf{True}, & \text{если } I(g,m) \geq \alpha; \\ \mathbf{False} & \text{в противоположном случае.} \end{cases}$$

На втором шаге используется методический комплекс классического АФП для вывода из $(G, M, I^{(\alpha)})$ четких понятий и построения их решетки.

Содержанием третьего шага является трансформация каждого полученного четкого понятия (X, Y) в нечеткое (X_f, Y). Объем X_f есть нечеткое множество над универсумом G. Оценка истинности $X_f(g)$ принадлежности объекта $g \in G$ к X_f определяется степенью, в которой он обладает всеми свойствами в содержании Y, а точнее - оценкой совместной истинности (пересечения) нечетких суждений «объекту $g \in G$ присуще свойство $m \in Y$» для всех свойств Y. Обычно предлагается оценивать эту степень принадлежности min-конъюнкцией:

$$X_f(g) = \begin{cases} \min_{m \in Y} I(g,m), & \text{если } g \in X; \\ 0 & \text{в противоположном случае.} \end{cases}$$

Нетрудно убедиться, что между такими нечеткими понятиями сохранится тот же частичный порядок, что и между четкими понятиями, получаемыми на промежуточном шаге метода одностороннего порога. Если дополнительно потребовать $\forall g \in G{:}\ Y = \varnothing \rightarrow X_f(g) = 1$, получаем, что сконструированные описанным способом нечеткие понятия - бикластеры (X_f, Y) с четким содержанием и нечетким объемом - образуют полную решетку нечетких понятий.

3.2. Метод нечеткого замыкания

Метод формирует нечеткие понятия с нечеткими объемами и с нечеткими содержаниями. В рассмотрение вводится алгебраическая структура, называемая полной резидуальной решеткой («решеткой с делением»), $\boldsymbol{L} = \langle L, \wedge, \vee, \otimes, \rightarrow, 0, 1\rangle$ такая что:

- $\langle L, \wedge, \vee, 0, 1\rangle$ - полная решётка с наименьшим элементом 0 и наибольшим элементом 1;
- $\langle L, \otimes, 1\rangle$ - коммутативный моноид;
- нечёткая конъюнкция $\otimes$ и нечёткая импликация $\rightarrow$ удовлетворяют условию сопряженности $x \otimes y \leq z \leftrightarrow x \leq y \rightarrow z$. Авторы [23] используют операторы Лукасевича

$$x \otimes y = \max(x + y - 1, 0),$$
$$x \rightarrow y = \min(1 - x + y, 1).$$

Оценивая эту формализацию, вслед за [24] отметим, что здесь как и в других подходах, основанных на нечетком замыкании (см., например, [13, 17]), возникают определенные проблемы интерпретации решеток нечетких понятий, которые зависят от используемой алгебраической структуры, т.к. вводимые в анализ алгебраические операции слабо связаны со смыслом и прагматикой приложений.

Продолжая обзор метода нечеткого замыкания, обозначим через L^Z множество всех нечетких множеств над универсумом Z.

Для нечетких множеств $X \in L^G$ и $Y \in L^M$ нечеткие множества $X\!\uparrow \in L^M$ и $Y\!\downarrow \in L^G$ определяются как

$$X\!\uparrow(m) = \wedge_{g \in G}(X(g) \rightarrow I(g, m)),$$
$$Y\!\downarrow(g) = \wedge_{m \in M}(Y(m) \rightarrow I(g, m)).$$

$X\uparrow(m)$ указывает степень истинности того, что свойство m характеризует все объекты в нечетком множестве X. Аналогично, $Y\downarrow(g)$ свидетельствует об истинности того, что объекту g присущи все свойства в нечетком множестве Y.

Пара $(X, Y) \in L^G \times L^M$ является нечетким формальным понятием, если $X\uparrow = Y$ и $Y\downarrow = X$. Множество всех нечетких формальных понятий, извлекаемых из нечеткого ФК, частично упорядочено по включению нечетких объемов (или, равно, нечетких содержаний) и образует полную решетку нечетких понятий.

Согласно [23] обнаружение всех нечетких понятий сводится к вычислению всех неподвижных точек определенного оператора нечеткого замыкания. В нечетком ФК составной оператор $\uparrow\downarrow$: $L^G \to L^G$ является оператором нечеткого замыкания в G, а $\downarrow\uparrow$: $L^M \to L^M$ - оператор нечеткого замыкания в M. Неподвижные точки операторов $\uparrow\downarrow$ и $\downarrow\uparrow$ определяют соответственно объемы и содержания нечетких формальных понятий.

4. Учет ограничений существования свойств

Очевидно (см. раздел 2), что для четких формальных понятий, извлекаемых из данного нечеткого ФК, «естественным» критерием учета ОСС является нормальность наборов свойств, определяющих содержание построенных понятий [10-12]. При выводе из нечеткого ФК нечетких формальных понятий этот подход нуждается в расширении. У нечеткого понятия содержание может быть нечетким множеством, которое прямо несопоставимо с четкими нормальными множествами, определяемыми ОСС. Поэтому в рассматриваемом случае следует опереться на более общее условие учета ОСС («фундаментальный» критерий): денотат (объект), порождаемый согласно любому нечеткому понятию данного нечеткого ФК, должен характеризоваться нормальным множеством свойств.

4.1. Ситуация при использовании метода одностороннего порога

Нетрудно видеть, что для учета ОСС в методе одностороннего порога достаточно при α-аппроксимации нечеткого соответствия I вместо стандартного α-сечения использовать метод рационального α-сечения [12, 25]. В этом случае содержания выводимых нечетких формальных понятий будут четкими нормальными подмножествами множества измеряемых свойств M, и «естественный» критерий учет ОСС оказывается удовлетворен.

4.2. Нормализация формального контекста для применения метода нечеткого замыкания

Метод построения нечетких понятий с использованием оператора нечеткого замыкания не использует пороговые значения, а содержания выводимых понятий являются нечеткими подмножествами – определенными элементами множества L^M. Поэтому из-за того, что требование выполнения ОСС формулируется на языке обычных множеств (объектам предметной области могут быть присущи лишь нормальные подмножества измеримых свойств M), возникает необходимость отыскания связи между элементами булеана 2^M и множества L^M.

Известно, что в случае $L = [0, 1]$ такую связь устанавливает теорема о декомпозиции нечетких множеств. В наших обозначениях для каждого $g \in G$ имеем

$$\{g\}\uparrow = \cup_{\alpha \in [0, 1]} \alpha\cdot(\{g\}\uparrow)_\alpha, \quad (1)$$

где $\{g\}\uparrow$ - нечеткое множество свойств объекта g; α - пороговое значение, $\alpha \in [0, 1]$, $(\{g\}\uparrow)_\alpha$ - четкое множество свойств объекта g уровня α.

Теперь требование выполнения «фундаментального» критерия учета ОСС можно применить к правой части (1): ОСС будут удовлетворены, если все множества $(\{g\}\uparrow)_\alpha$ являются нормальными. Заметим, что в реальном ФК для каждого $g \in G$ количество различных «слагаемых» в (1) конечно.

Резюмируя, для учета ОСС можно предложить следующий результативный метод предварительной обработки исходного нечеткого ФК (G, M, I) для построения решетки нечетких формальных понятий:

- согласно имеющимся ОСС выявляются все нормальные подмножества множества измеряемых свойств M;
- соответствие I подвергается нормализации: у каждого $g \in G$ в качестве исходного нечеткого множества свойств принимается правая часть (1) за исключением «слагаемых», где множество свойств $(\{g\}\uparrow)_\alpha$ не является нормальным.

5. Заключение

Выполненное исследование позволило указать пути совмещения дедуктивных достижений НАФП в построении решеток нечетких формальных понятий с пониманием роли второго – априорного - аспекта общей гипотетико-дедуктивной природы АФП. Конкретно предложены способы учета экзистенциальных отношений на множестве наблюдаемых и/или измеренных свойств, обеспечивающие вывод корректных нечетких понятий.

Перспективной задачей являет исследование влияния ограничений существования свойств на качественные и количественные характеристики решеток нечетких понятий в зависимости от параметров этих ограничений и параметров исходных формальных контекстов.

Практическая ценность результатов исследования состоит в повышении адекватности приложений НАФП. В частности, он используется при построении нечетких формальных онтологий, важная роль которых высоко оценена в обширной зарубежной (см., например, упоминавшиеся работы [20, 22]) и отечественной литературе (см., например, [26]).

6. Литература

[1] Ganter, B. Formal Concept Analysis. Mathematical foundations / B. Ganter, R. Wille – Berlin-Heidelberg: Springer-Verlag, 1999. – 290 p.

[2] Ignatov, D.I. Introduction to Formal Concept Analysis and Its Applications in Information Retrieval and Related Filds / D.I. Ignatov, P. Braslavski, N. Karpov, M. Worring, Y. Volkovich, D.I. Ignatov // Information Retrieval. Revised Selected Papers 8th Russian Summer School – Springer International Publishing, 2015. – P. 42-141.

[3] Formal Concept Analysis Homepage [Electronic resource]. – Access mode: http://www.upriss.org.uk/fca (25.12.2019).

[4] Гетманова, А.Д. Логика. Углубленный курс: учебное пособие – М.: Кнорус, 2016. – 192 с.

[5] Понятие [Electronic resource]. – Access mode: https://ru.wikipedia.org/wiki/%D0%9F%D0%BE%D0%BD%D1%8F%D1%82%D0%B8%D0%B5 (25.12.2019).

[6] Pollandt, S. Fuzzy-Begriffe: Formale Begriffsanalyse unscharfer Daten – Berlin-Heidelberg: Springer-Verlag, 1997. – 145 P.

[7] Belohlavek, R. What is a Fuzzy Concept Lattice? / R. Belohlavek, V. Vychodil // CEUR Workshop Proceedings. – 2005. – Vol. 162. – P. 34-45.

[8] Смирнов, С.В. Двухсоставность феномена информации и анализ данных (с примерами из когнитивного анализа) // Сборник трудов III Международной конференции и молодежной школы «Информационные технологии и нанотехнологии» (ИТНТ) – Самара: Новая техника, 2017. – С. 1846-1849.

[9] Самойлов, Д.Е. Фрактальность ограничений сосуществования свойств в задачах машинного обучения / Д.Е. Самойлов, В.А. Семенова, С.В. Смирнов // Сборник трудов IV международной конференции и молодежной школы «Информационные технологии и нанотехнологии» ИТНТ – Самара: Новая техника, 2018. – С. 2512-2518.

[10] Lammari, N. Building and maintaining ontologies: a set of algorithms / N. Lammari, E. Metais // Data & Knowledge Engineering. – 2004. – Vol. 48(2). – P. 155-176.

[11] Пронина, В.А. Использование отношений между атрибутами для построения онтологии предметной области / В.А. Пронина, Л.Б. Шипилина // Проблемы управления. – 2009. – Т. 1. – С. 27-32.

[12] Офицеров, В.П. Метод альфа-сечения нестрогих формальных контекстов в анализе формальных понятий / В.П. Офицеров, В.С. Смирнов, С.В. Смирнов // Труды XVI международной конференции «Проблемы управления и моделирования в сложных системах» – Самара: СамНЦ РАН, 2014. – С. 228-244.

[13] Belohlavek, R. What is a Fuzzy Concept Lattice? II // Rough Sets, Fuzzy Sets, Data Mining and Granular Computing – Berlin-Heidelberg: Springer, 2011. – P. 19-26.
[14] Belohlavek, R. Factor Analysis of Incidence Data via Novel Decomposition of Matrices / R. Belohlavek, V. Vychodil – Berlin-Heidelberg: Springer, 2009. – P. 83-97.
[15] Cross, V. Comparing Two Approaches to Creating Fuzzy Concept Lattices / V. Cross, M. Kandasamy, W. Yi // Proc. of the North American Fuzzy Information Processing Society, 2011. – P. 233-238.
[16] Glodeanu, C.V. Attribute Exploration in a Fuzzy Setting // CEUR Workshop Proceedings. – 2012. – Vol. 876. – P. 114-128.
[17] Панкратьева, В.В. О взаимосвязи между различными подходами к анализу нечетких формальных понятий / В.В. Панкратьева // Интеллектуальные системы. Теория и приложения. – 2016. – Т. 20, № 3. – С. 158-163.
[18] Boffa, S. Unifying fuzzy concept lattice construction methods / S. Boffa, C. De Maio, A. Di Nola, G. Fenza, A.R. Ferraioli, V. Loia // Proc. of 2016 IEEE International Conference on Fuzzy Systems, 2016. – P. 209-216.
[19] Zhang, Z. Constructing L-fuzzy concept lattices without fuzzy Galois closure operation // Fuzzy Sets and Systems. – 2018. – Vol. 333. – P. 71-86. DOI: 10.1016/j.fss.2017.05.002.
[20] Tho, Q.T. Automatic Fuzzy Ontology Generation for the Semantic Web / Q.T. Tho, S.C. Hui, A.C.M. Fong, T.H. Cao // IEEE Transactions on Knowledge and Data Engineering. – 2006. – Vol. 18(6). – P. 842-856.
[21] Yang, K.M. Fuzzy Concept Mining based on Formal Concept Analysis / K.M. Yang, E.H. Kim, S.H. Hwang, S.H. Choi // Int. J. of Computers. – 2008. – Vol. 2(3). – P. 279-290.
[22] De Maio, C. Towards Automatic Fuzzy Ontology Generation / C. De Maio, L.V. Fenza, S. Senatore // Proc. of IEEE International Conference on Fuzzy Systems, 2009. – P. 1044-1049.
[23] Belohlavek, R. Computing the lattice of all fixpoints of a fuzzy closure operator / R. Belohlavek, B. De Baets, B. Outrata, J. Vychodil // IEEE Trans. on Fuzzy systems. – 2010. – Vol. 18(3). DOI: 10.1109/TFUZZ.2010.2041006.
[24] Wolff, K.E. Position Paper: Pragmatics in Fuzzy Theory // Rough Sets, Fuzzy Sets, Data Mining and Granular Computing – Berlin-Heidelberg: Springer, 2011. – P. 135-138.
[25] Самойлов, Д.Е. Эвристический алгоритм дефаззификации исходного контекста в анализе формальных понятий / Д.Е. Самойлов, В.А. Семенова, С.В. Смирнов // Сборник трудов V международной конференции и молодежной школы «Информационные технологии и нанотехнологии» (ИТНТ) – Самара: Новая техника, 2019. Т. 4. – С. 9-16.
[26] Ярушкина, Н.Г. Применение способа интеграции нечетких временных рядов и нечетких онтологий в задачах диагностики технических систем / Н.Г. Ярушкина, В.С. Мошкин, Г.Р. Ишмуратова, И.А. Андреев, И.А. Мошкина // Онтология проектирования. – 2018. – Т. 8, №4(30). – С. 594-604. DOI: 10.18287/2223-9537-2018-8-4-594-604.

Properties Existence Constraints in Fuzzy Formal Concept Analysis

A.E. Samoilov[1,2], S.V. Smirnov[2]

[1]Samara National Research University, Moskovskoe Shosse 34A, Samara, Russia, 443086
[2]Institute for the Control of Complex Systems of RAS - branch of the Samara Federal Scientific Center of RAS, Sadovaya str. 61, Samara, Russia 443020

Abstract. The field of research is the creating of fuzzy concept lattices based on fuzzy data "objects - properties". Our contribution is the account of existential relations on the set of observed and / or measured properties, i.e. "properties existence constraints". Two most well-known approaches to c creating of fuzzy concepts lattices are considered: the one-sided threshold and fuzzy closure methods. It is shown that for the more popular one-sided threshold method, potential violations of properties existence constraints in the concept lattice are countered by the rational threshold cut method, previously developed to elicitation crisp formal concepts from fuzzy initial data. However, this way is fundamentally unacceptable for the fuzzy closure method. For this case, the idea of special preliminary processing of the initial data is put forward - the "normalization" of the fuzzy set of properties for each object in the training sample. The practical importance of the study is to increase an adequacy of Fuzzy Formal Concept Analysis.