

Эвристический алгоритм дефаззификации исходного контекста в анализе формальных понятий

Д.Е. Самойлов^{1,2}, В.А. Семенова^{2,3}, С.В. Смирнов²

¹Самарский национальный исследовательский университет им. академика С.П. Королева, Московское шоссе 34А, Самара, Россия, 443086

²Институт проблем управления сложными системами РАН, Садовая 61, Самара, Россия, 443020

³Самарский государственный технический университет, Молодогвардейская 244, Самара, Россия, 443100

Аннотация. Областью интереса данной работы является проблема извлечения из эмпирических данных решетки формальных понятий, которая может служить основой формальной онтологии исследуемой предметной области. Исходному эмпирическому материалу, т.е. данным многомерных наблюдений и экспериментов, как правило, свойственна неполнота и противоречивость, обусловленная реалиями накопления эмпирической информации. Это приводит к тому, что нужный для построения решетки формальный контекст – соответствие «объекты-свойства» - предварительно может быть представлен лишь в рамках какой-либо многозначной логики. Его требуется аппроксимировать в бинарной логике, поскольку эффективные методы вывода формальных понятий разработаны лишь для однозначных (бинарных) формальных контекстов. Точное решение этой задачи, учитывающее ограничения существования свойств у объектов исследуемой предметной области, во-первых, затруднительно, а, во-вторых, в некотором смысле неадекватно ожиданиям субъекта, исследующего предметную область. Предлагается использовать для дефаззификации исходного формального контекста эвристику, суть которой состоит в локализации задачи аппроксимации исходного «мягкого» контекста на уровне каждой группы зависимых свойств каждого объекта обучающей выборки. Полная модель ограничений существования свойств, формируется как множество иерархий групп зависимых свойств, что предопределяет рекурсивный и многопроходный характер разработанного алгоритма дефаззификации.

1. Введение

Стандартно протоколы наблюдений и экспериментов оформляются как таблицы «объекты-свойства» (ТОС) [1, 2]. Чаще всего ТОС содержит результаты измерения значений некоторого множества свойств у набора объектов, оказавшихся в поле зрения исследователя. Подобные ТОС именуется *многозначными формальными контекстами* (ФК) задач анализа данных [3]. Вместе с тем, весьма популярным методом «раскопки» данных «объекты-свойства» стал *анализ формальных понятий* (АФП) [3-7], который имеет дело лишь с бинарными (*однозначными*) ФК:

$$(G^*, M, I), \quad (1)$$

где $G^* = \{g_i\}_{i=1, \dots, r}$, $r = |G^*| \geq 1$ - множество наблюдавшихся объектов: $G^* \subseteq G$, где G – всё гипотетически мыслимое множество объектов исследуемой предметной области (ПрО);

$M = \{m_j\}_{j=1,\dots,s}$, $s = |M| \geq 1$ - набор измеряемых у объектов свойств; $I \subseteq G^* \times M$ - бинарное отношение, описываемое матрицей инцидентности «объекты-свойства», каждый элемент которой суть оценка истинности базового семантического суждения о ПрО (БСС) $b_{ij} = \langle \text{«объекту } g_i \in G^* \text{ присуще свойство } m_j \in M \rangle$, т.е. $I = (\|b_{ij}\|)_{i=1,\dots,r; j=1,\dots,s}$, $\|b_{ij}\| \in \{\text{Истина, Ложь}\}$.

Основной целью АФП является извлечение из табличных данных кластеров, называемых *формальными понятиями*. Частично упорядоченная совокупность всех формальных понятий называется решеткой понятий, которая может рассматриваться как «скелет» *формальной онтологии* исследуемой ПрО [8-10].

Однозначный ФК может быть получен некоторой, определяемой целями исследования ПрО *грануляцией* информации стандартного многозначного ФК с помощью так называемых *концептуальных шкал* [11, 12]. Этот инструмент генерации гранулярности аналогичен лингвистическим переменным, введенным Заде [13]. Нечеткое концептуальное шкалирование [14] приводит к ФК вида (1) с нечетким отношением I . Несколько иной подход к концептуальному шкалированию, формирующему нечеткое отношение I , исследовался в [15].

В работах [16-18] генезис «мягкого» (в частности, нечеткого) ФК рассмотрен с более общих позиций и в целом связан с неполнотой и противоречивостью первичной информации об исследуемой ПрО. При этом установлено, что нечеткая логика для описания «мягкого» соответствия «объекты-свойства» менее адекватна, чем более изощренные многозначные логики. В [17, 18] для формирования оценок истинности БСС предложено использовать *векторную логику* [19], конкретно - логику V^{TF} , которая может рассматриваться как простейшее обобщение нечеткой логики Заде.

Конструктивное использование «мягкого» ФК основано на его α -аппроксимации, т.е. замене в (1) нечеткого (или *нестрогого* в случае использования векторных логик) отношения I его α -сечением $I^{(\alpha)}$ [12, 14-18]. К сожалению, в общем случае однозначное соответствие $I^{(\alpha)}$ в задаче анализа формальных понятий оказывается *некорректным*, поскольку процедура стандартного α -сечения не учитывает *экзистенциальные зависимости* между измеряемыми свойствами: *несовместимость* и *обусловленность* свойств у объектов исследуемой ПрО [18, 20, 21]. Исчерпывающая модель *системы измеряемых свойств* с ограничениями их существования (СИС) обоснована в [22]. Данное сообщение посвящено разработке алгоритма *рациональной α -аппроксимации нестрого ФК*, учитывающего выявленную сложную организацию СИС в задачах АФП.

2. Оценка истинности суждений о свойствах объекта

В логике V^{TF} истинность БСС b_{ij} оцениваться *вектором* $\langle \text{Истина, Ложь} \rangle$ (рисунок 1):

$$\|b_{ij}\| = \langle b^+_{ij}, b^-_{ij} \rangle; b^+_{ij}, b^-_{ij} \in [0, 1]; (b^+_{ij} + b^-_{ij} \equiv 1) = \text{Ложь}.$$

Каждый объект обучающей выборки $g_i \in G^*$, $i = 1, \dots, r$, характеризуется в нестрогом ФК множеством векторов истинности $\{\|b_{ij}\|\}_{j=1,\dots,s}$ (рисунок 2). V^{TF} -порог доверия $\alpha = \langle \alpha^+, \alpha^- \rangle$ определяет в пространстве существования векторов истинности две области, попадание вектора $\|b_{ij}\|$ в одну из которых интерпретируется как *ложность*, а в другую как *истинность* соответствующего БСС (рисунок 3). С другой стороны, в указанном пространстве для каждого вектора $\|b_{ij}\|$ определены две области, попадание порога доверия α в одну из которых означает *ложность*, а в другую - *истинность* БСС b_{ij} (рисунок 4).

3. Паттерны групп сопряженных свойств

Согласно [22] СИС – это множество иерархий *групп сопряженных свойств* (ГСС), которое формируется как продукт гипотезирования субъекта о структуре понятийного описания исследуемой ПрО. Каждая иерархия ГСС образуется путем *замещения* в супергруппе одного или нескольких измеряемых свойств субгруппой таких свойств; высшую супергруппу в иерархии вложенности ГСС составляет *протосвойство*.

Существуют три *базовых* паттерна субгрупп, замещающих в супергруппе *одно* свойство:

- группа из двух измеряемых свойств, в которой одно свойство *обуславливает* другое;
- группа из $n > 1$ *несовместимых* измеряемых свойств;

- группа из $n > 1$ взаимно обусловленных свойств.

Характер сопряженности свойств в таких ГСС можно наглядно отразить в пространстве существования векторов истинности, связывая его точки надлежащими дугами (рисунок 5).

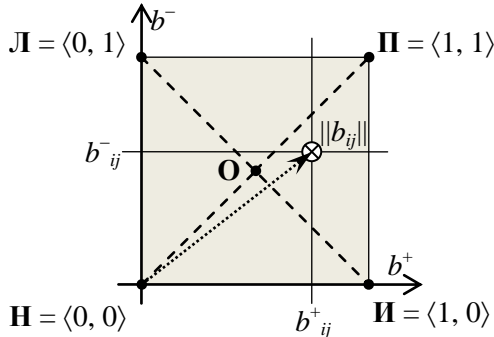


Рисунок 1. Представление вектора истинности $\|b_{ij}\| = \langle b^+_{ij}, b^-_{ij} \rangle$ и истинностных констант V^{TF} -логики «Истина», «Ложь», «Неопределенность», «Противоречие», «Недоопределенность» (И, Л, Н, П, О).

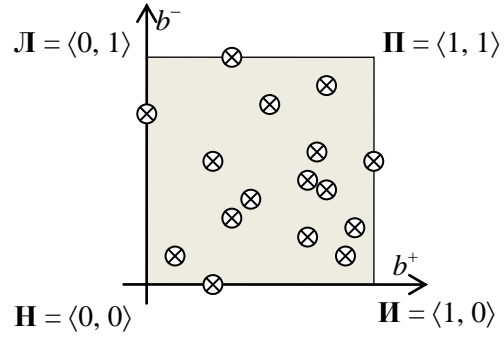


Рисунок 2. Множество векторов истинности $\{\|b_{ij}\|\}_{j=1, \dots, s}$, характеризующих в нестрогом формальном контексте объект обучающей выборки g_i .

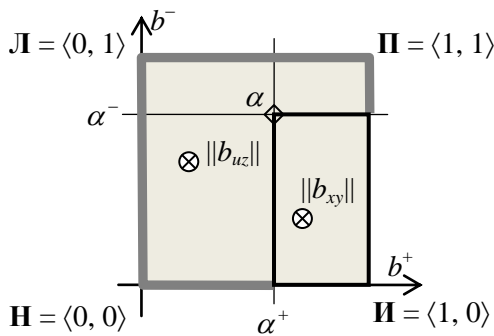

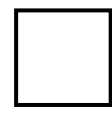


Рисунок 3. Определение множества уровня α (или α -сечения; $\alpha = \langle \alpha^+, \alpha^- \rangle$) нестрогого множества через отношение логического порядка для векторов истинности в V^{TF} -логике: $\|b_{xy}\| > \alpha$, если $b^+_{xy} \geq \alpha^+, b^-_{xy} \leq \alpha^-$.

-  - область, попадание в которую вектора $\|b_{ij}\|$ интерпретируется как *ложность* соответствующего БСС;
-  - область, попадание в которую вектора $\|b_{ij}\|$ интерпретируется как *истинность* соответствующего БСС.

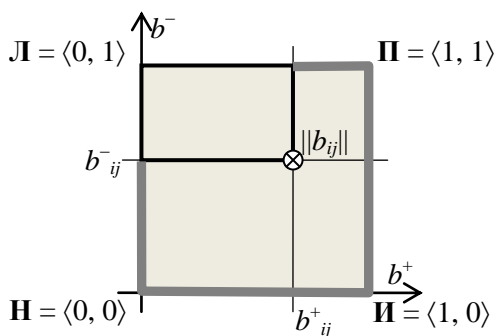
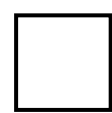



Рисунок 4. Области выбора порога доверия к исходным данным $\alpha = \langle \alpha^+, \alpha^- \rangle$, определяющие приближенную оценку истинности в шкале {Истина, Ложь} БСС, векторная оценка истинности которого равна $\langle b^+_{ij}, b^-_{ij} \rangle$.

-  - область, при выборе в которой порога доверия $\alpha = \langle \alpha^+, \alpha^- \rangle$, БСС с векторной оценкой истинности $\langle b^+_{ij}, b^-_{ij} \rangle$ признается *истинным*;
-  - область, при выборе в которой порога доверия $\alpha = \langle \alpha^+, \alpha^- \rangle$, БСС с векторной оценкой истинности $\langle b^+_{ij}, b^-_{ij} \rangle$ признается *ложным*.

Аналогично может быть представлены два *сложных* паттерна субгрупп, замещающих по фиксированным правилам *несколько* свойств в супергруппе (*только* в супергруппах, где все свойства взаимно обусловлены):

- группа свойств с несколькими обусловленностями (рисунок ба);
- группа свойств с несколькими несовместимостями и обусловленностями (рисунок бб).

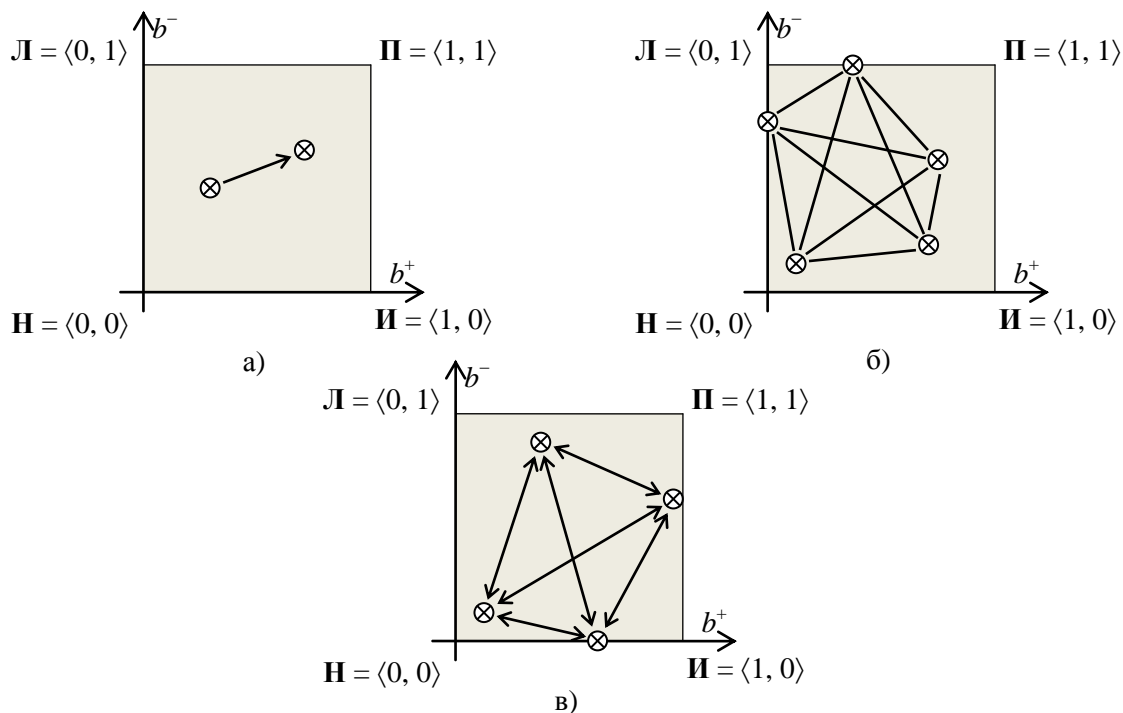


Рисунок 5. Примеры совмещенных диаграмм сопряженности свойств и оценок истинности БСС, соответствующих этим свойствам, для базовых паттернов групп сопряженных свойств: а) группа, где одно свойство обуславливает другое; б) группа несовместимых свойств; в) группа взаимно обусловленных свойств.

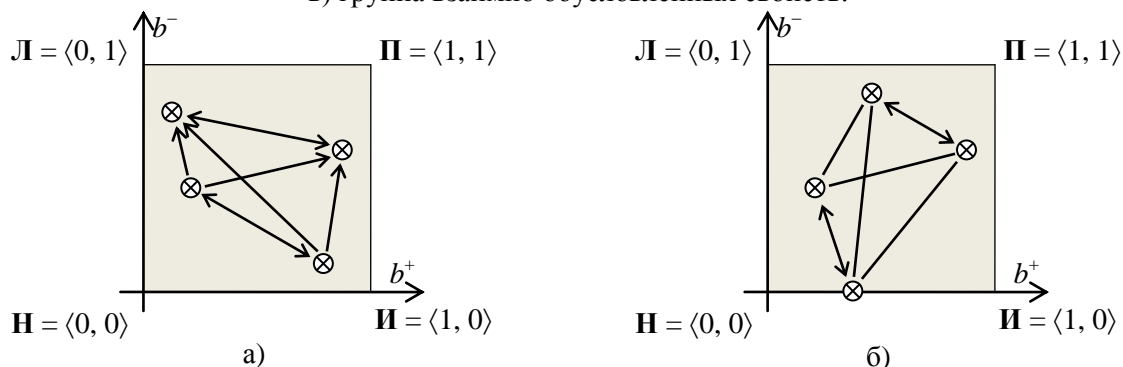


Рисунок 6. Примеры совмещенных диаграмм сопряженности свойств и оценок истинности БСС, соответствующих этим свойствам, для сложных паттернов групп сопряженных свойств: а) группа свойств с несколькими обусловленностями; б) группа свойств с несколькими несовместимостями и обусловленностями.

4. Рациональное альфа-сечение нестроого формального контекста

Итак, с целью применения эффективных методов АФП для извлечения формальных понятий из нестроого ФК необходимо обеспечить его корректную α -аппроксимацию при наличии ограничений существования свойств (ОСС).

4.1. Поиск области допустимых значений порога доверия к исходным данным

Формально поставленную задачу можно свести к построению одноместного предиката « α -сечение корректно» с векторным аргументом $\alpha = \langle \alpha^+, \alpha^- \rangle$, $\alpha^+, \alpha^- \in [0, 1]$, где условие подтверждения истинности каждого эмпирического БСС b_{ij}

$$b^+_{ij} \geq \alpha^+ \wedge b^-_{ij} \leq \alpha^-$$

(или, напротив, необходимой ложности этого БСС) должно быть совмещено с выполнением всех актуальных ОСС. И далее найти область (не исключено, что она окажется пуста) существования порогов доверия α , доставляющих такому предикату значение **Истина**.

Разумеется, в общем случае построить такой предикат и выявить указанную область весьма непросто; пример построения такого предиката для случая, когда ОСС вызваны несовместимостью некоторой части измеряемых свойств, можно найти в [23]. Однако, даже допуская возможность такого решения, связать субъекта, исследующего ПрО, необходимостью выбирать порог только из *ограниченной области* весьма *непрактично*. Например, при таком подходе для субъекта могут не оправдываться интуитивные ожидания от смягчения или ужесточения порога доверия к данным, представляющим ПрО [23].

4.2. Эвристический подход

Взамен описанному поиску предлагается следующая *эвристика*:

- субъект свободен в выборе порога (выбор α -порога произволен);
- определяемый порогом α и, в общем случае, недопустимый состав свойств *каждого* объекта искомого однозначного ФК (этому соответствует выделение на рисунке 2 части оценок $\|b_{ij}\|$ по правилу, иллюстрированному рисунком 3) следует последовательно сокращать за счет поочередного *отсечения* свойств, нарушающих ОСС;
- механизм отсечения состоит в *локальном ужесточении* порога доверия в пределах *каждой* ГСС в СИС, когда основным критерием выбора измеряемого свойства для отсечения служит *минимальное ужесточение* (например, в прямоугольной метрике) выбранного субъектом порога доверия.

Предлагаемый эвристический метод получения корректной однозначной аппроксимации нестроого ФК *результативен*.

Действительно, метод реализуется для каждого объекта обучающей выборки *сепарательно*, и на некотором шаге возникнет одно из следующих условий:

- либо множество свойств объекта начинает удовлетворять ОСС (заметим, что пустое множество свойств таким ограничением удовлетворяет, однако тогда объект с таким «набором» свойств придется квалифицировать как неопознанный);
- либо будет констатировано, что в исходных «мягких» оценках истинности БСС имеется *неустранимое противоречие*: относительно объекта имеются БСС, истинные в классической двузначной логике, но нарушающие ОСС.

4.3. Алгоритм дефаззификации нестроого формального контекста

Сложность алгоритмизации предложенной эвристики обусловливается иерархической структурной организацией СИС (см. секцию 3). Следующие утверждения определяют конструкцию алгоритма дефаззификации нестроого ФК.

Утверждение 1. Для отсечения отдельных свойств, нарушающих ОСС, каждая ГСС должна иметь порог отсечения *всех* составляющих ее свойств, или «ложь-порог» $\beta = \langle \beta^+, \beta^- \rangle$. Согласно рисунку 4 ложь-порог ГСС с точностью до бесконечно малой равен оценке истинности БСС для входящего в данную группу свойства, если эта оценка занимает крайнее «юго-восточное» положение в единичном квадрате векторных оценок истинности. В общем случае это положение занимают несколько оценок истинности БСС для составляющих данную группу свойств, которые вместе с разделяющими их точками пространства векторов истинности образуют множество Парето.

Утверждение 2. Порог доверия, устанавливаемый в супергруппе вначале субъектом, а затем, возможно, неоднократно подвергаемый ужесточению, «касается» всех ее свойств, включая замещенные субгруппами. Следовательно, этот «истина-порог» должен быть применен ко всем субгруппам рассматриваемой супергруппы независимо от «глубины» положения в иерархии ГСС.

Утверждение 3. Новый истина-порог ГСС с точностью до бесконечно малой устанавливается равным оценке истинности БСС, соответствующего *отсекаемому* свойству (см. рисунок 4). В

общем случае критерий минимального ужесточения приводит к формированию Парето-множества оценок истинности БСС, соответствующих свойствам, образующим множество кандидатов на отсечение.

Утверждение 4. Однозначный выбор в Парето-множествах векторов истинности обеспечивают два вспомогательных критерия их сравнения: *мера достоверности* и *мера определенности* [19], которые для вектора $\langle b^+, b^- \rangle$ целесообразно вычислять по формулам $\mu_d = b^+ - b^-$ и $\mu_o = \max(b^+, b^-)$ соответственно.

Утверждение 5. Вид *паттерна* структурной организации ГСС (рисунки 5, 6) определяет прием поиска экстремальных точек в пространстве оценок истинности БСС, соответствующих свойствам, которые включены в анализируемую ГСС.

В главном цикле алгоритма для каждого объекта обучающей выборки g_i последовательно выполняются две рекурсивные процедуры:

- первая вычисляет ложь-порог для каждой ГСС;
- вторая опирается на результаты первой и реализует собственно используемую эвристику: для каждой ГСС отыскивается значение локального истина-порога, которое либо идентифицирует часть членов данной ГСС как свойства объекта g_i , либо исключает ГСС из состава свойств объекта g_i .

5. Заключение

Разработанный алгоритм дефаззификации исходного контекста призван учесть в задачах анализа формальных понятий ранее не принимавшуюся во внимание сложную организацию системы измеряемых свойств исследуемой предметной области.

Алгоритм ориентирован на однозначную аппроксимацию нестрогих формальных контекстов, поэтому актуально создание его версии для обработки широко используемых на практике (несмотря на меньшую адекватность) нечетких соответствий «объекты-свойства».

Перспективной задачей является исследование эффективности предложенного алгоритма в зависимости от параметров, характеризующих систему измеряемых свойств исследуемой предметной области и исходный «мягкий» контекст задачи анализа формальных понятий.

6. Литература

- [1] Барсегян, А.А. Анализ данных и процессов / А.А. Барсегян, М.С. Куприянов, И.И. Холод, М.Д. Тесс, С.И. Елизаров. – СПб.: БХВ-Петербург, 2009. – 512 с.
- [2] Загоруйко, Н.Г. Когнитивный анализ знаний / Н.Г. Загоруйко. – Новосибирск: Академическое изд-во «Гео», 2013. – 186 с.
- [3] Ganter, B. Formal Concept Analysis. Mathematical foundations / B. Ganter, R. Wille. – Springer-Verlag Berlin Heidelberg, 1999. – 290 p.
- [4] Formal Concept Analysis Homepage [Электронный ресурс]. – Режим доступа: <http://www.upriss.org.uk/fca/fca.html> (19.11.2018).
- [5] Ganter, B. Conceptual Exploration / B. Ganter, S. Obiedkov. – Springer-Verlag Berlin Heidelberg, 2016. – 315 p.
- [6] Ignatov, D.I. Introduction to Formal Concept Analysis and Its Applications in Information Retrieval and Related Fields / D.I. Ignatov // Information Retrieval. Revised Selected Papers 8th Russian Summer School. – Springer International Publishing, 2015. – P. 42-141.
- [7] Carpineto, C. Concept Data Analysis: Theory and Applications / C. Carpineto, G. Romano. – Wiley, 2004. – 220 p.
- [8] Obitko, M. Ontology Design with Formal Concept Analysis / M. Obitko, V. Snasel, J. Smid // Proc. of the CLA International Workshop on Concept Lattices and their Applications. – TU of Ostrava, Dept. of Computer Science, 2004. – P. 111-119.
- [9] Смирнов, С.В. Онтологический анализ предметных областей моделирования // Известия Самарского научного центра РАН. – 2001. – Т. 3, № 1. – С. 62-70.

- [10] Godin, R. *Ontology Design with Formal Concept Analysis* / R. Godin, H. Mili, G.W. Mineau, R. Missaoui, A. Arfi, T.T. Chau // *Theory and Application of Object Systems (TAPOS)*. – 1998. – Vol 4(2). – P. 117-134.
- [11] Ganter, B. *Conceptual scaling* / B. Ganter, R. Wille // *Applications of Combinatorics and Graph Theory to the Biological and Social Sciences*. – New York Springer-Verlag, 1989. – P. 139-167.
- [12] Wolff, K.E. *Conceptual Interpretation of Fuzzy Theory* / K.E. Wolff // *6th European Congress on Intelligent Techniques and Soft Computing*. – Aachen, 1998. – P. 555-562.
- [13] Zadeh, L.A. *The concept of a linguistic variable and its application to approximate reasoning* // *Information Science*. – 1975. – Vol. 8. – P. 199-249 (Part I); Vol. 8. – P. 301-357 (Part II); Vol. 9. – P. 43-80 (Part III).
- [14] Belohlavek, R. *Scaling, Granulation, and Fuzzy Attributes in Formal Concept Analysis* / R. Belohlavek, J. Konecny // *The IEEE International Conference on Fuzzy Systems*. – P. 918-923.
- [15] Pollandt, S. *Fuzzy-Begriffe: Formale Begriffsanalyse unscharfer Daten*. – Springer-Verlag Berlin Heidelberg, 1996. – 153 p.
- [16] Смирнов, С.В. *Нечеткие формальные контексты при построении онтологий на основе анализа формальных понятий: происхождение и использование* // *Знания – Онтологии – Теории: Труды всероссийской конф. с международным участием*. – Новосибирск: Институт математики СО РАН, 2007. – Т. 2. – С. 17-25.
- [17] Смирнов, С.В. *Многозначная и нечеткая логики в онтологическом анализе данных* // *Информационные технологии и системы: Труды Третьей междунар. науч. конф. ИТиС*. – Челябинск: Изд-во Челябинского гос. ун-та, 2014. – С. 90-91.
- [18] Самойлов, Д.Е. *Анализ неполных данных в задачах построения формальных онтологий* / Д.Е. Самойлов, В.А. Семенова, С.В. Смирнов // *Онтология проектирования*. – 2016. – Т. 6, № 3(21). – С. 317-339. DOI: 10.18287/2223-9537-2016-6-3-317-339.
- [19] Аршинский, Л.В. *Векторные логики: основания, концепции, модели*. – Иркутск: Иркут. гос. ун-т, 2007. – 228 с.
- [20] Пронина, В.А. *Использование отношений между атрибутами для построения онтологии предметной области* / В.А. Пронина, Л.Б. Шипилина // *Проблемы управления*. – 2009. – № 1. – С. 27-32.
- [21] Lammari, N. *Building and maintaining ontologies: a set of algorithms* / N. Lammari, E. Metais // *Data & Knowledge Engineering*. – 2004. – Vol. 48(2). – P. 155-176.
- [22] Самойлов, Д.Е. *Структура системы измеряемых свойств у объектов многомерного наблюдения и экспериментов* / Д.Е. Самойлов, В.А. Семенова, С.В. Смирнов // *Вестник Самарского государственного технического университета. Серия «Технические науки»*. – 2018. – № 3(59). – С. 56-71.
- [23] Офицеров, В.П. *Метод альфа-сечения нестрогих формальных контекстов в анализе формальных понятий* / В.П. Офицеров, В.С. Смирнов, С.В. Смирнов // *Проблемы управления и моделирования в сложных системах: Труды XVI международной конференции*. – Самара: СамНЦ РАН, 2014. – С. 228-244.

Heuristic algorithm for defuzzification of initial context in formal concept analysis

D.E. Samoilov^{1,2}, V.A. Semenova^{2,3}, S.V. Smirnov²

¹Samara National Research University, Moskovskoe Shosse 34A, Samara, Russia, 443086

²Institute for Control of Complex Systems of RAS, Sadovaya str. 61, Samara, Russia 443020

³Samara State Technical University, Molodogvardeyskaya str. 244, Samara, Russia, 443100

Abstract. The research field is the problem of extracting from the initial empirical material the formal concept lattice, which can serve as the basis of the formal ontology of the studied subject domain. The initial empirical material, i.e. the data of multidimensional observations and experiments, is characterized by incompleteness and inconsistency, conditioned by realities of empirical information accumulation. This leads to the fact that required for lattice building formal context can be previously presented only within the framework of some multi-valued logic. It needs to be approximated in binary logic, since effective methods for derivation of formal concepts are developed only for unambiguous (binary) formal contexts. The exact solution of this problem, considering the properties existence constraints of objects in the studied subject domain, is difficult and in a certain sense is inadequate to expectations of subject exploring the subject domain. For defuzzification of the initial formal context heuristic was proposed, idea of which is to localize the approximation task of "soft" context within every group of dependent properties of each object of learning sample. The model reflecting such restrictions is formed as hierarchy of groups of dependent properties, which predetermines the recursive and multi-pass nature of the developed defuzzification algorithm.