

Методы и приложения искусственного интеллекта

УДК 519.677

DOI: 10.25729/ESI.2023.30.2.001

Эффективность редукции ограничений существования свойств в задаче идентификации признаков объекта

Смирнов Сергей Викторович, Семенова Валентина Андреевна

Самарский федеральный исследовательский центр РАН,
Институт проблем управления сложными системами РАН,
Россия, Самара, smirnov@iccs.ru

Аннотация. В статье рассматривается проблема учета знания об ограничениях существования в часто встречающейся на практике задаче выявления присущих объекту свойств (признаков, или атрибутов) из априори определяемого набора измеряемых у объекта свойств в случае неполноты и противоречивости результатов измерений. Ставится вопрос о возможности повысить производительность разработанного для решения этой задачи авторского методического обеспечения за счет сокращения размерности описания ограничений существования свойств, измеряемых у объекта. Естественной моделью ограничений существования является множество сущностей (например, признаков объекта) с двумя заданными на нём бинарными отношениями сосуществования – несовместимостью и обусловленностью; размерность такой модели определяется количеством сущностей и мощностью отмеченных экзистенциальных отношений. Показано, что желаемая редукция модели ограничений существования измеряемых у объекта свойств осуществима на основе выделения во множестве измеряемых свойств классов эквивалентности и определения расширенных бинарных отношений несовместимости и обусловленности на множестве выделенных классов. Количественная оценка эффективности редукции ограничений существования выполнена путём машинных имитационных статистических экспериментов. Самостоятельный интерес представляет примененная тактика имитационного моделирования, обусловленная многосвязной структурой ограничений существования, а также многоплановое использование для оценки надлежащего количества статистических испытаний неравенства Хёфдинга, устанавливающего экспоненциально убывающие оценки вероятности отклонений сумм независимых случайных величин от среднего этих сумм. Результаты проведенного исследования подтвердили значимость сокращения размерности модели ограничений существования и выявили характер зависимости получаемого выигрыша от параметров, определяющих конфигурацию указанных ограничений.

Ключевые слова: атрибутивная идентификация объекта, ограничения существования, классы эквивалентности, имитационное статистическое моделирование

Цитирование: Смирнов С.В. Эффективность редукции ограничений существования свойств в задаче идентификации признаков объекта / С.В. Смирнов, В.А. Семенова // Информационные и математические технологии в науке и управлении. – 2023. – № 2 (30). – С. 05-13. DOI: 10.25729/ESI.2023.30.2.001.

Введение. Построение семантической модели объекта, определяющей его признаки, или, иначе говоря, состав присущих объекту свойств, актуальна в онтологическом анализе эмпирических данных [1], распознавании образов [2], структурном синтезе проектно-конструкторских решений (см., например, [3, 4]) и др. В подобных задачах идентифицируемый состав свойств объекта является подмножеством всех измерявшихся (в широком смысле, включая, в частности, использование экспертиз) у объекта свойств. При этом такая атрибутивная идентификация объекта существенно усложняется, если необходим учет априори заданных ограничений существования свойств (ОСС) [5-7] – бинарных отношений несовместимости и обусловленности на множестве измеряемых свойств.

Предложенная в [8, 9] методика выявления всех допустимых наборов признаков объекта включает логико-комбинаторный метод выбора «лучшего» из них на основе анализа достоверности эмпирических свидетельств о присущих объекту свойствах. В основе метода лежит ограниченный перебор всех пар свойств с обусловленностью и несовместимостью (О- и Н-

пар). Указанный там же способ увеличения производительности метода связан с редукцией описания ОСС [10]. Данная статья посвящена исследованию эффективности этой редукции методом имитационного статистического моделирования [11, 12].

1. Ограничения существования. Пусть g - идентифицируемый объект, M – множество измеряемых свойств, $P_g(x)$ – одноместный предикат «объекту g присуще свойство x », $x \in M$. Тогда можно дать следующие формальные определения экзистенциональных отношений, образующих ОСС:

- несовместимость $E: M \times M \rightarrow \{\mathbf{True}, \mathbf{False}\}$, $E(x, y) \leftrightarrow P_g(x) \rightarrow \neg P_g(y)$, $x, y \in M$;
- обусловленность $C: M \times M \rightarrow \{\mathbf{True}, \mathbf{False}\}$, $C(x, y) \leftrightarrow P_g(x) \rightarrow P_g(y)$, $x, y \in M$.

Наглядно ОСС представляет граф, в котором вершины представляют измеряемые у объекта свойства, ребра – несовместимость пары связываемых свойств, дуги – обусловленность свойства, находящегося в стоке дуги, свойством, представленным её вершиной-истоком (рис. 1).

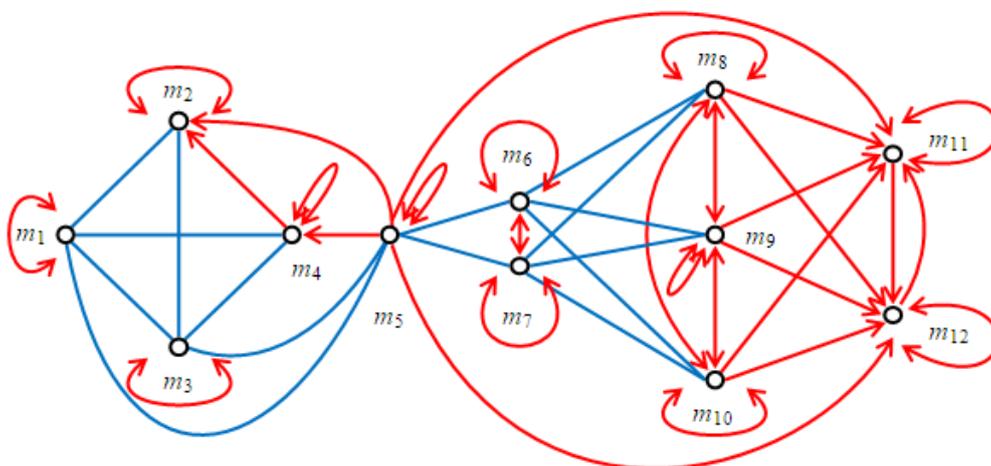


Рис. 1. Пример описания ограничений существования: вершина графа – свойство; ребро – несовместимость, дуга – обусловленность пары свойств; двунаправленная дуга – обозначение 2-х встречных дуг и естественной самообусловленности свойства

Обусловленность рефлексивна, несимметрична и транзитивна, а несовместимость антирефлексивна, симметрична и нетранзитивна, но характеризуется так называемой «транзитивностью относительно обусловленности» (далее — E -транзитивность), что означает $\forall x, y, z \in M: C(x, y) \wedge E(y, z) \rightarrow E(x, z)$.

Наличие ОСС ограничивает поиск признаков описания объекта лишь среди подмножеств $X \subseteq M$, которые замкнуты и совместимы [5-7], т.е. $\forall x \in X: (\exists y \in M, x \neq y: C(x, y)) \rightarrow y \in X$ и $\forall x \in X: (\exists y \in M, x \neq y: E(x, y)) \rightarrow y \notin X$.

2. Редукция описания ограничений существования. Легко видеть, что отношение обусловленности $C(x, y)$ индуцирует на множестве измеряемых свойств M бинарное отношение взаимообусловленности $MC: M \times M \rightarrow \{\mathbf{True}, \mathbf{False}\}$, $MC(x, y) \leftrightarrow C(x, y) \wedge C(y, x)$, которое включает пары взаимообусловленных свойств (ВЗО-пары).

Поскольку MC рефлексивно, симметрично и транзитивно, то оно разбивает множество M на классы эквивалентности [13, 14], которыми являются группы взаимообусловленных свойств (ВЗО-групп).

Опираясь на транзитивность отношения C и E -транзитивность отношения E , нетрудно определить и доказать существование расширенных отношений обусловленности C^* и несовместимости E^* на множестве классов эквивалентности M^* , т.е. на множестве ВЗО-групп измеряемых свойств [6, 10].

Предложенный в [8] метод атрибутивной идентификации объекта по неполным и противоречивым эмпирическим данным реализуется практически одинаково для описания ОСС в формах (M, E, C) и (M^*, E^*, C^*) . Полезный эффект перехода к (M^*, E^*, C^*) выражается в сокращении размерности задачи. Механизм такой редукции иллюстрирует рис. 2:

- уменьшается количество сущностей, состоящих в расширенных отношениях обусловленности и несовместимости, – вместо свойств рассматриваются их ВЗО-группы: $|M^*| \leq |M|$ (см. рис. 2а);
- уменьшается количество пар, состоящих в экзистенциальных отношениях – вместо О- и Н-пар свойств анализируются возникающие О- и Н-пары ВЗО-групп $|E^*| \leq |E|$, $|C^*| \leq |C|$ (см. рис. 2б и 2в).

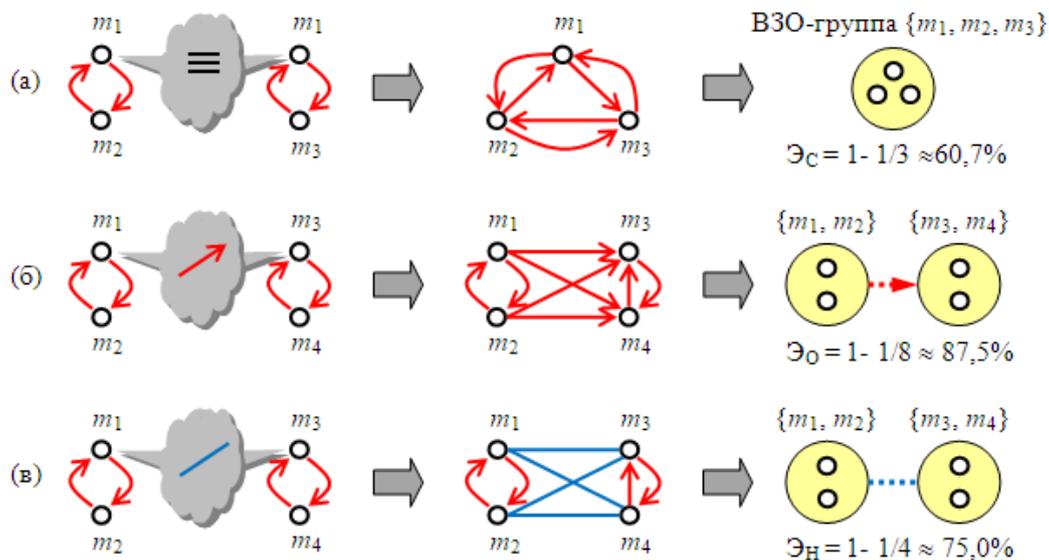


Рис. 2. Примеры редукции ограничений существования свойств и достигаемый при этом выигрыш (самообусловленность свойств опущена, расширенные экзистенциальные связи показаны пунктирной линией)

На рис. 3 приведено редукцированное описание ОСС, исходный вид которого представлен на рис. 2. Очевидно, что масштабы редукции определяются характеристиками исходной модели ОСС и, прежде всего, мощностью множества ВЗО-пар в СИС.

3. Организация вычислительных экспериментов. Количественно эффект редукции ОСС оценивался в имитационных статистических экспериментах, когда при заданной величине $|M|$ структура ОСС рандомизировалась при варьировании долями ВЗО-, О-, Н-пар, и исследовался выигрыш от редукции ОСС – среднее относительное уменьшение количества сущностей $\Theta_C = 1 - |M^*|/|M|$ (см. рис. 2а) и их экзистенциальных связей – обусловленностей $\Theta_O = 1 - |C^*|/|C|$ (см. рис. 2б) и несовместимостей $\Theta_H = 1 - |E^*|/|E|$ (см. рис. 2в).

Важный в подобных исследованиях вопрос о надлежащем количестве статистических испытаний предпочтительно решать на основе неравенства Хёфдинга [15, 16]:

$$Pr\{|S_N - E(S_N)| \geq N\varepsilon\} \leq 2 \exp\left(-\frac{2N^2\varepsilon^2}{\sum_{i=1}^N (b_i - a_i)^2}\right), \quad (1)$$

где S_N – сумма независимых случайных величин X_1, X_2, \dots, X_N , таких, что $a_i \leq X_i \leq b_i$ почти наверняка, $E(S_N)$ – математическое ожидание величины S_N , $\varepsilon > 0$. Неравенство Хёфдинга даёт

экспоненциально убывающие оценки вероятности отклонений сумм независимых случайных величин от среднего этих сумм. Эти оценки точнее, чем аналогичные, полученные с использованием первых или вторых моментов (например, неравенства Маркова, Чебышёва), обеспечивающие лишь степенной закон убывания указанных отклонений.

Учитывая, что оцениваемые показатели \mathcal{E}_C , \mathcal{E}_O , \mathcal{E}_N лежат в диапазоне $[0, 1]$ и рассчитываются, как средние значения наблюдавшихся в испытаниях относительных сокращений соответствующих характеристик, редуцированных ОСС, на основании неравенства (1) нетрудно получить

$$N \geq \frac{1}{2\varepsilon^2} \ln \frac{2}{\delta}, \quad \forall \delta, \varepsilon \in (0, 1),$$

т.е. утверждать, что при количестве статистических испытаний N получаемые оценки показателей \mathcal{E}_C , \mathcal{E}_O , \mathcal{E}_N с вероятностью не меньше $1 - \delta$ будут аппроксимировать их действительные значения с абсолютной погрешностью ε .

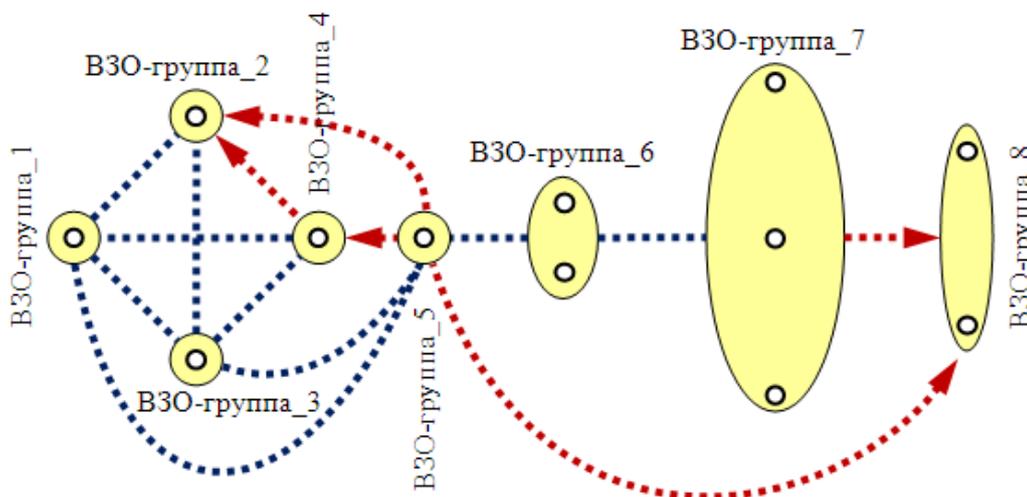


Рис. 3. Пример [8] описания ограничений существования свойств в форме расширенных отношений обусловленности и несовместимости на множестве групп взаимообусловленных измеряемых свойств (описание обозначений см. на рис. 2)

В табл. 1 приведены варианты соответствующего расчета, демонстрирующие чувствительность минимально необходимого количества испытаний к изменению параметров достоверности и погрешности оценок.

Таблица 1. Количество статистических испытаний для оценки исследуемых показателей

Достоверность оценок $1 - \delta$	Абсолютная погрешность оценок ε	Нижняя граница минимально необходимого количества испытаний	Использовавшееся в экспериментах количество испытаний
0,90	0,05	$200 \cdot \ln 20 \approx 599,1$	600
0,90	0,01	$5000 \cdot \ln 20 \approx 14978,7$	15000
0,95	0,01	$5000 \cdot \ln 40 \approx 18444,4$	18450

Особенность имитации ОСС в исходной форме (M, E, C) заключается в естественных ограничениях на количество одновременно присутствующих ВЗО-, О- и Н-пар - $N_{\text{ВЗО-п}}$, $N_{\text{О-п}} = |C|$, $N_{\text{Н-п}} = |E|$ соответственно:

$$2N_{\text{ВЗО-п}} + N_{\text{О-п}} + 2N_{\text{Н-п}} \leq |M| \cdot (|M| - 1),$$

где $N_{\text{ВЗО-п}}, N_{\text{Н-п}} \in \{0, 1, 2, \dots, |M| \cdot (|M| - 1) / 2\}$, $N_{\text{О-п}} \in \{0, 1, 2, \dots, |M| \cdot (|M| - 1)\}$.

Кроме того, некоторые конфигурации ОСС (здесь и далее «конфигурация» – это сочетание параметров $|M|, N_{\text{ВЗО-п}}, N_{\text{О-п}}, N_{\text{Н-п}}$) в указанном смысле могут не существовать. Например,

легко проверить, что при $|M| = 3$ имеем $N_{\text{ВЗО-п}} \in \{0, 1, 3\}$, т.к. при попытке построить ОСС с $N_{\text{ВЗО-п}} = 2$ вследствие транзитивности обусловленности количество ВЗО-пар закономерно становится равным 3. Аналогично при $|M| = 3$ добавление 5-й обусловленности в ОСС, где существование свойств определяется 4-мя О-парами, приводит к образованию ОСС с 6-ю О-парами, т.е. при указанном количестве измеряемых свойств имеем $N_{\text{О-п}} \in \{0, 1, 2, 3, 4, 6\}$. В присутствии обусловленностей подобные ограничения возникают и для Н-пар.

Отмеченная феноменология формирования ОСС потребовала выработки следующей, представляющей самостоятельный интерес, тактики имитационного статистического моделирования:

- имитация ОСС с заданными количествами ВЗО-, О- и Н-пар одновременно рассматривается как испытание Бернулли, успех которого констатирует существование требуемой конфигурации ОСС. Если N таких испытаний неудачны (здесь можно воспользоваться тем же значением необходимого числа испытаний, которое определено для «целевого» эксперимента с задачей наблюдения эффекта редукции ОСС, ибо результат испытания Бернулли можно интерпретировать как реализацию дискретной случайной величины с двумя возможными значениями – «0» и «1», которые, равно как и оцениваемые показатели эффективности редукции, принадлежат ограниченному диапазону $[0, 1]$), то эксперимент с ОСС, определяемой данными параметрами прекращается;
- при возникновении в результате имитации ОСС с заданным количеством ВЗО-, О- и Н-пар закономерного перехода ОСС в «состояние», характеризуемые иными значениями $N_{\text{ВЗО-п}}$, $N_{\text{О-п}}$, $N_{\text{Н-п}}$, разумно не считать эту фазу вычислительного эксперимента неудачей, которая заставляет повторить попытку имитации ОСС с заданными параметрами, а продолжить имитационный эксперимент (т.е. редуцировать «возникшую» ОСС с регистрацией соответствующих эффектов), накапливая статистику для ставшей актуальной конфигурации ОСС, и лишь затем обратиться к новой попытке имитации ОСС с целевыми параметрами. Фактически это означает «параллельное» выполнение экспериментов для всех актуальных конфигураций ОСС, требуя подготовки и использования в общем случае $N_{\text{ВЗО-п}} \times N_{\text{О-п}} \times N_{\text{Н-п}}$ организующих и регистрирующих структур, что, тем не менее, компенсируется сокращением вычислительных затрат на имитационное моделирование.

4. Результаты вычислительных экспериментов. Эффект уменьшения количества сущностей в редуцированных ОСС возникает целиком благодаря наличию в исходном описании ОСС ВЗО-пар и усиливается вследствие их пересечений (см. рис. 2). Поэтому эксперименты для количественной оценки этого эффекта, т.е. определения показателя \mathcal{E}_c , выполнялись для ОСС, характеризуемых параметрами $N_{\text{ВЗО-п}} \in \{0, 1, 2, \dots, |M| \cdot (|M| - 1) / 2\}$, $N_{\text{О-п}} = 2N_{\text{ВЗО-п}}$ и $N_{\text{Н-п}} = 0$. При этом априори очевидно, что при $N_{\text{ВЗО-п}} = 0$ эффект отсутствует ($\mathcal{E}_c = 0$), а при $N_{\text{ВЗО-п}} = |M| \cdot (|M| - 1) / 2$ максимален: $\mathcal{E}_c^{\max} = 1 - 1/|M|$. Таким образом, область существования значений \mathcal{E}_c на графике зависимости этого показателя от доли $q_{\text{ВЗО-п}} = N_{\text{ВЗО-п}} / (|M| \cdot (|M| - 1) / 2)$ в значительной степени определена, что иллюстрирует рис. 4а. Результат конкретного исследования для $|M| = 20$ представлен на рисунке 4б.

Исследования показали, что с увеличением в ОСС доли ВЗО-пар наблюдается почти экспоненциальный с насыщением рост выигрыша, оцениваемого показателем \mathcal{E}_c . Подобная затухающая экспонента характерна для отклика систем с отрицательной обратной связью, хотя в области насыщения наблюдается множество «разрывов» – отсутствие конфигураций ОСС с определенным количеством ВЗО-пар – и существенный разброс величины выигрыша. Такая картина выигрыша \mathcal{E}_c принципиально не меняется при варьировании количеством измеряемых свойств M .

На рисунке 5а приведен пример экспериментально установленного среднего относительного уменьшения количества обусловленностей \mathcal{E}_O в редуцированных ОСС для $|M| = 20$ измеряемых свойств в зависимости от количества ВЗО-пар $N_{ВЗО-п}$ и доли О-пар $q_{O-п} = N_{O-п} / (|M| \cdot (|M| - 1))$ при $N_{H-п} = 0$ в исходном представлении ОСС.

Выигрыш \mathcal{E}_O сначала значительный (и тем выше, чем больше ВЗО-пар в исходном представлении ОСС), но довольно быстро снижается. Это можно объяснить тем, что при фиксированном $N_{ВЗО-п}$ рост $q_{O-п}$ происходит за счет О-пар, не пересекающихся с существующими ВЗО-парами (пересечение вызвало бы образование дополнительных ВЗО-пар, а их количество фиксировано) и не образующих новых ВЗО-пар, а обусловленности, определяемые такими «чистыми» О-парами, сохраняются в редуцированных ОСС в неизменном виде, сокращая выигрыш.

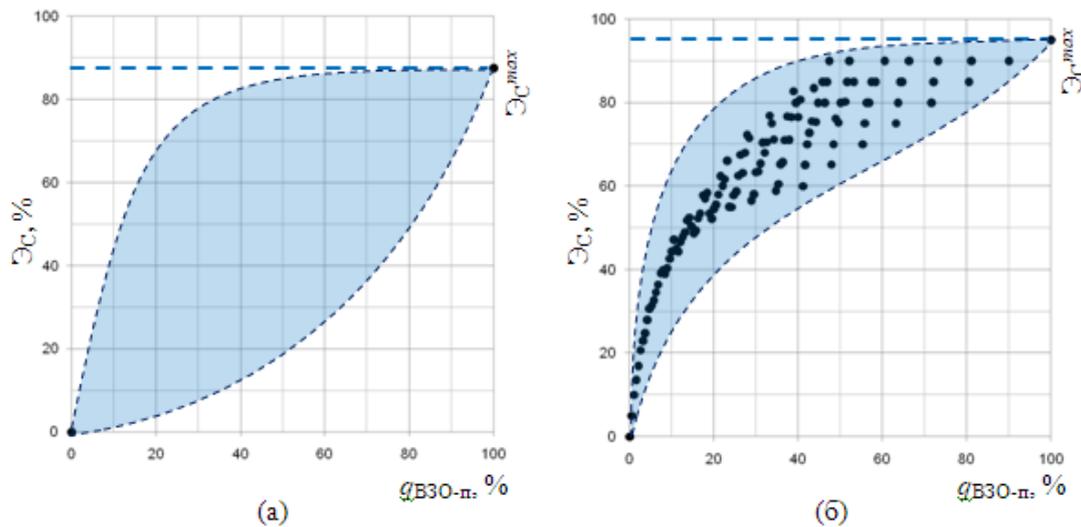


Рис. 4. Зависимость показателя \mathcal{E}_C – среднего относительного уменьшения количества сущностей в редуцированном представлении ограничений существования свойств – в зависимости от доли ВЗО-пар $q_{ВЗО-п}$ в исходном представлении этих ограничений: (а) – область существования значений \mathcal{E}_C ; (б) – пример для $|M| = 20$ измеряемых свойств

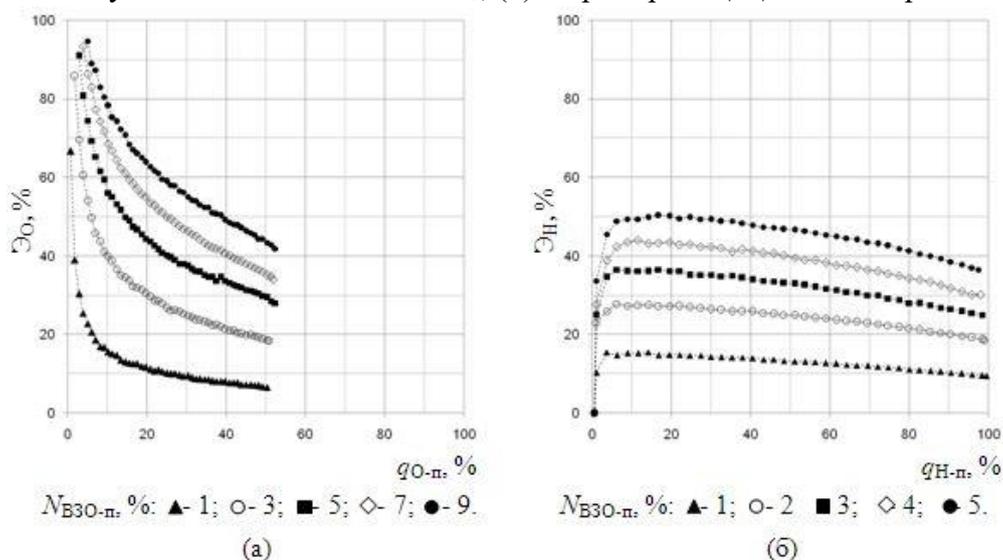


Рис. 5. Пример зависимости показателей \mathcal{E}_O (а) и \mathcal{E}_H (б) – соответственно среднего относительного уменьшения количества обусловленностей и несовместимостей в редуцированном представлении ограничений существования свойств – в зависимости от количества ВЗО-пар $N_{ВЗО-п}$ и долей $q_{O-п}$ (для \mathcal{E}_O), $q_{H-п}$ (для \mathcal{E}_H) в исходном представлении этих ограничений для $|M| = 20$ измеряемых свойств

Правую границу области существования графиков \mathcal{E}_O определяет максимум количества «чистых» О-пар $-|M| \cdot (|M| - 1)/2 - N_{ВЗО-п}$; левую – количество О-пар, образующих исходное количество ВЗО-пар в ОСС при отсутствии «чистых» О-пар; доля таких О-пар составляет $2N_{ВЗО-п} / (|M| \cdot (|M| - 1))$.

Рисунок 5б демонстрирует пример среднего относительного уменьшения количества несовместимостей \mathcal{E}_H в редуцированных ОСС для $|M| = 20$ в зависимости от количества ВЗО-пар $N_{ВЗО-п}$ и доли Н-пар $q_{H-п} = N_{H-п} / (|M| \cdot (|M| - 1)/2)$ при $N_{O-п} = 0$ в исходном представлении ОСС. Наблюдаются следующие характерные черты зависимостей такого вида:

- рост выигрыша \mathcal{E}_H при увеличении $N_{ВЗО-п}$;
- нулевой выигрыш в начале всех графиков \mathcal{E}_H – при $N_{H-п} = 1$, – поскольку существование единственной Н-пары в ОСС возможно лишь при отсутствии её пересечения с существующими в ОСС ВЗО-парами. Несовместимость, определяемая такими Н-парами, сохраняется в редуцированных ОСС в неизменном виде;
- быстрое достижение с ростом $q_{H-п}$ максимальной величины выигрыша, который медленно и почти линейно уменьшается до достижения величиной $N_{H-п}$ своего максимума $|M| \cdot (|M| - 1)/2 - N_{ВЗО-п}$, определяющего правую границу области существования графиков \mathcal{E}_H .

Заключение. В статье применительно к задаче атрибутивной идентификации объекта на основе данных многомерных наблюдений и измерений теоретически и экспериментально подтверждена целесообразность редуцирования описания ограничений существования свойств путём выделения во множестве измеряемых у объекта свойств классов эквивалентности и определения расширенных экзистенциальных отношений на множестве выделенных классов.

Представленное исследование позволило количественно оценить потенциал снижения размерности у описания ограничений существования свойств и выявить характер зависимости получаемого выигрыша от параметров, определяющих конфигурацию указанных ограничений.

Полученные знания являются вкладом в методологию семантического моделирования объектов и предметных областей, имеющую широкое приложение в системных исследованиях, проектировании и интеллектуальном анализе данных.

Благодарности. Работа выполнена при финансовой поддержке Минобрнауки России, код научной темы FMRW-2022-0030.

Список источников

1. Смирнов С.В. Онтологический анализ предметных областей моделирования / С.В. Смирнов // Известия Самарского научного центра РАН, 2001. – Т. 3. – № 1. – С. 62-70.
2. Закревский А.Д. Логика распознавания / А.Д. Закревский. – Мн.: Наука и техника, 1988. – 118 с.
3. Половинкин А.И. Основы инженерного творчества / А.И. Половинкин. – М.: Машиностроение, 1988. – 386 с.
4. Егер С.М. Основы автоматизированного проектирования самолетов / С.М. Егер, Н.К. Лисейцев, О.С. Самойлович. – М.: Машиностроение, 1986. – 232 с.
5. Lammari N. Building and maintaining ontologies: a set of algorithms. *Data & Knowledge Engineering*, 2004, vol. 48 (2), pp. 155-176.
6. 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.
7. Пронина В.А. Построение онтологии предметной области с нормализацией контекста в методе формальных параметров / В.А. Пронина, Л.Б. Шипилина // 11-я нац. конф. по искусственному интеллекту с международным участием (28 сентября – 3 октября 2008 г., Дубна, Россия): труды. – М: URSS, 2008. – Т. 1. – С. 197-206.
8. Semenova V., Smirnov S. Revealing attributes of an object based on incomplete and inconsistent empirical data. VIII Int. Conf. on Information Technology and Nanotechnology (ITNT). IEEE Xplore, 2022, pp. 1-5, DOI: 10.1109/ITNT55410.2022.9848601.

9. Семенова В.А. Эвристика и численный метод нормализации эмпирического V^{TF} -контекста в онтологическом анализе данных / В.А. Семенова // Информационные и математические технологии в науке и управлении, 2021. – № 2 (22). – С. 61-69. – DOI:10.38028/ESI.2021.22.2.006.
10. Семенова В.А. Алгоритмизация формирования и прагматической трансформации ограничений существования свойств предметной области / В.А. Семенова, С.В. Смирнов // Онтология проектирования, 2020. – Т. 10. – № 3. – С. 361-379. – DOI: 10.18287/2223-9537-2020-10-3-361-379.
11. Кобелев Н.Б. Имитационное моделирование / Н.Б. Кобелев, В.В. Девятков, В.А. Половников – 2-е изд. – М: КУРС, 2021. – 352 с.
12. Михайлов Г.А. Статистическое моделирование. Методы Монте-Карло / Г.А. Михайлов, А.В. Войтишек – М.: Изд-во Юрайт, 2022. – 324 с.
13. Каазик Ю.Я. Математический словарь / Ю.Я. Каазик. – М.: Физматлит, 2007. – 334 с.
14. Лелонд О.В. Дискретная математика / О.В. Лелонд, М.А. Тренина. – Тольятти: Изд-во Тольяттинского гос. ун-та, 2018. – 93 с.
15. Hoeffding W. Probability inequalities for sums of bounded random variables. J. of the American Statistical Association, 1963, v. 58, pp. 13-30 (In: The Collected Works of Wassily Hoeffding. N.I. Fisher et al. (eds.) - Springer, New York, 1994, ch. 26, p. 409-426, DOI: 10.1007/978-1-4612-0865-5_26).
16. Петров В.В. Предельные теоремы для сумм независимых случайных величин / В.В. Петров. - Изд. 2. – М.: URSS, 2022. – 320 с.

Смирнов Сергей Викторович. Д.т.н., доцент, главный научный сотрудник Института проблем управления сложными системами Самарского федерального исследовательского центра РАН, профессор Поволжского государственного университета телекоммуникаций и информатики. Область научных интересов – анализ и моделирование сложных систем. AuthorID: 17628, SPIN: 8039-1421, ORCID: 0000-0002-3332-5261. smirnov@iccs.ru. Россия, Самара, ул. Садовая, 61.

Семенова Валентина Андреевна. Младший научный сотрудник Института проблем управления сложными системами Самарского федерального исследовательского центра РАН. Область научных интересов – анализ и моделирование сложных систем. AuthorID: 1063366, ORCID: 0000-0002-0557-3890. queenbfjr@gmail.com. Россия, Самара, ул. Садовая, 61.

UDC 519.677

DOI: 10.25729/ESI.2023.30.2.001

Efficiency of properties existence constraints reduction in the problem of object attributive identification

Sergey V. Smirnov, Valentina A. Semenova

Samara Federal Research Scientific Center RAS, Institute for the Control of Complex Systems RAS, Russia, Samara, *smirnov@iccs.ru*

Abstract. The article deals with the problem of taking into account the existence constraints in the widespread task of identifying the properties inherent in an object (or features, or attributes) from an a priori determined set of properties measured in an object in the case of incompleteness and inconsistency of measurement results. The question is raised about the possibility of increasing the productivity of the author's methodological support developed for solving this problem by reducing the description dimension of the existence constraints of properties measured in an object. A natural model of existence constraints is a set of entities (for example, attributes of an object) with two binary coexistence relations defined on it - incompatibility and conditionality; the dimension of such a model is determined by the number of entities and the quantity of the noted existential relations. It is shown that the desired reduction of the existence constraints model of properties measured in an object is feasible based on the identification of equivalence classes in the set of measured properties and the definition of extended binary incompatibility and conditionality relations on the set of discovered classes. Quantitative assessment of the existence constraints reduction effectiveness was made by computer simulation statistical experiments. Of independent interest is the simulation tactics used, due to the multiply connected structure of existence constraints, as well as the multifaceted use of Höfding's inequality to estimate the appropriate number of statistical tests, which establishes exponentially decreasing estimates of the deviations probability of sums of independent random variables from the average of these sums. The results of the study confirmed the significance of dimension reducing of the existence constraints model and revealed the dependence nature of the gain obtained on the parameters that determine the configuration of these constraints.

Keywords: attributive object identification, existence constraints, equivalence classes, statistical simulation

Acknowledgements: This research was funded by Ministry of Science and Higher Education of the Russian Federation, research topic code FMRW-2022-0030.

References

1. Smirnov S.V. Ontologicheskii analiz predmetnykh oblastey modelirovaniya [Ontological analysis of knowledge domains of modeling]. *Izvestiya Samarskogo nauchnogo tsentra RAN* [Bulletin of the Samara Scientific Center of RAS], 2001, vol. 3, no. 1, pp. 62-70.
2. Zakrevskiy A.D. Logika raspoznavaniya [Recognition logic]. Mn., Nauka i tekhnika [Minsk. Science and technology], 1988. 118 p.
3. Polovinkin A.I. Osnovy inzhenernogo tvorchestva [Fundamentals of engineering creativity]. M.: Mashinostroyeniye [Moscow. Mechanical engineering], 1988, 386 p.
4. Yeger S.M., Lisejcev N.K., Samojlovich O.S. Osnovy avtomatizirovannogo proektirovaniya samolyotov [Fundamentals of Computer-Aided Aircraft Design]. M., Mashinostroyeniye [Moscow. Mechanical engineering], 1986, 232 p.
5. Lammari N., Metais E. Building and maintaining ontologies: a set of algorithms. *Data & Knowledge Engineering*, 2004, vol. 48 (2), pp. 155-176.
6. Pronina V.A., Shipilina L.B. Ispol'zovanie otnosheniy mezhdu atributami dlya postroeniya ontologii predmetnoy oblasti [Using the relationships between attributes to build domain ontology]. *Problemy upravleniya* [Control problems], 2009, no. 1, pp. 27-32.
7. Pronina V.A., Shipilina L.B. Postroenie ontologii predmetnoy oblasti s normalizatsiej konteksta v metode formal'nykh parametrov [Building an ontology of a knowledge domain with context normalization in the method of formal parameters]. 11-ya nacional'naya konferenciya po iskustvennomu intellektu s mezhdunarodnym uchastiem (28 sentyabrya – 3 oktyabrya, 2008, Dubna, Rossiya). *Trudy* [Proc. of 11th National conf. on artificial intelligence with international participation (28 September – 3 October, 2008, Dubna, Russia)], Moscow. URSS, 2008, vol. 1, pp. 197-206.
8. Semenova V., Smirnov S. Revealing attributes of an object based on incomplete and inconsistent empirical data. VIII Int. Conf. on Information Technology and Nanotechnology (ITNT). *IEEE Xplore*, 2022, pp. 1-5. DOI: 10.1109/ITNT55410.2022.9848601.
9. Semenova V.A. Evristika i chislennyi metod normalizatsii empiricheskogo V^{TF} -konteksta v ontologicheskom analize dannykh [Heuristics and numerical method for normalizing the empirical V^{TF} -context in ontological data analysis]. *Informatsionnyye i matematicheskiye tekhnologii v nauke i upravlenii* [Information and mathematical technologies in science and management], 2021, no. 2 (22), pp. 61-69. DOI:10.38028/ESI.2021.22.2.006.
10. Semenova V.A., Smirnov S.V. Algoritmizatsiya formirovaniya i pragmaticheskoy transformatsii ogranichenij suschestvovaniya svoystv predmetnoy oblasti [Algorithms for the formation and pragmatic transformation of existence constraints]. *Ontologiya proektirovaniya* [Ontology of Designing], 2020, vol. 10, no 3 (37), pp. 361-379. DOI: 10.18287/2223-9537-2020-10-3-361-379.
11. Kobelev N.B., Devyatkov V.V., Polovnikov V.A. Imitacionnoye modelirovaniye [Simulation], 2d ed. Moscow. KURS, 2021. 352 p.
12. Mihajlov G.A., Vojtishkek A.V. Statisticheskoye modelirovaniye. Metody Monte-Karlo [Statistical modeling. Monte Carlo Methods]. M. Izd-vo Yurayt [Moscow. Jurajt Publ.], 2022. 324 p.
13. Kaazik Ju.Ja. Matematicheskij slovar' [Math dictionary]. M. Fizmatlit [Moscow. Phizmatlit], 2007. 334 p.
14. Lelond O.V., Trenina M.A. Diskretnaya matematika [Discrete Math]. *Tol'yatti: Izd-vo Tol'yattinskogo gos. un-ta* [Togliatti. Togliatti State University Publ.], 2018. 93 p.
15. Hoeffding W. Probability inequalities for sums of bounded random variables. *J. of the American Statistical Association*, 1963, vol. 58, pp. 13–30 (In *The Collected Works of Wassily Hoeffding*, N.I. Fisher et al. (eds.). Springer, New York, 1994, ch. 26, pp. 409-426. DOI: 10.1007/978-1-4612-0865-5_26).
16. Petrov V.V. Predel'nyye teoremy dlya sum nezavisimyykh sluchajnykh velichin [Limit theorems for sums of independent random variables]. 2d ed. Moscow. URSS, 2022. 320 p.

Sergey V. Smirnov. D. Sc. (Tech.), Assoc. Chief Researcher at Institute for the Control of Complex Systems (Samara Federal Research Scientific Center RAS), professor at Povolzhskiy State University of Telecommunication and Informatics. Research interests – analysis and modeling of complex systems. Author ID: 17628, SPIN: 8039-1421, ORCID: 0000-0002-3332-5261. smirnov@iccs.ru. Russia, Samara, Sadovay st., 61.

Valentina A. Semenova. Junior Researcher at Institute for the Control of Complex Systems (Samara Federal Research Scientific Center RAS). Research interests – analysis and modeling of complex systems. Author ID: 1063366, ORCID: 0000-0002-0557-3890. queenbfjr@gmail.com. Russia, Samara, Sadovay st., 61.

Статья поступила в редакцию 27.03.2023; одобрена после рецензирования 15.05.2023; принята к публикации 15.06.2023.

The article was submitted 03/27/2023; approved after reviewing 05/15/2023; accepted for publication 06/15/2023.