
Секция 3 | МОДЕЛИРОВАНИЕ РАССУЖДЕНИЙ

УДК 004.83

doi: 10.15622/rcai.2025.016

О ВОЗМОЖНОСТИ ПОРОЖДЕНИЯ ОБОСНОВАННЫХ КАУЗАЛЬНЫХ ГИПОТЕЗ В ДСМ-МЕТОДЕ ДЛЯ ЗАДАЧИ НАСЛЕДСТВЕННОЙ БОЛЕЗНИ

Г.С. Вельмакин (*grigoryyii@gmail.com*)

Федеральный исследовательский центр
«Информатика и управление» РАН, Москва

В настоящей работе в ДСМ-методе, на примере обратного предиката сходства, будет построен предикат сходства с композицией отношений. Будет показано, что построенный предикат выразим в Граф-АФП – расширения анализа формальных понятий путём добавления отношений между объектами. В конце, посредством Граф-АФП, будет показана связь между предикатом сходства с композицией отношений и логикой описания. Всё вышеописанное позволит адекватно описать порождение каузальных гипотез в задаче изучения причин наследственной болезни в ДСМ-методе и Граф-АФП.

Ключевые слова: ДСМ-метод, предикат сходства с композицией отношений, Граф-АФП, логика описания.

Введение

Порождение обоснованных гипотез о причине той или иной наследственной болезни является актуальной задачей. Для её адекватного решения тем или иным методом необходимо, чтобы средствами используемого

метода можно было выразить композицию таких отношений как «родитель-ребёнок». ДСМ-метод [Финн, 2009], как метод, в котором происходит порождение (посредством индукции (п.п.в. I-го рода) и аналогии (п.п.в. II-го рода)) и обоснование (посредством абдукции) каузальных гипотез, мог бы быть использован для решения указанной задачи, однако в настоящий момент в нём отсутствуют средства, позволяющие схватить идею родословной ветви, как следствие, идею устойчивости паттерна «подобъект С есть причина множества свойств D» при наследовании, из-за чего, в частности, используя ДСМ-метод без отношений, в данной задаче, на основе такой БД, в которой есть два независимых родителя с исследуемой болезнью, но ни у их родителей, ни у их детей изучаемой болезни нет, как нет и у других людей из БД, посредством п.п.в. I-го рода мы сделаем грубый вывод, что болезнь наследственная. Настоящая работа призвана заполнить указанный пробел.

В первой части работы будет описана модельная задача наследования болезни. Для неё будет построен предикат сходства с композицией отношений ; его «объектами» являются множества упорядоченных троек $\langle C_1, C_2, C_3 \rangle$, а операцией сходства \odot является поэлементное сходство объектов, т. е. (для простоты восприятия мы оставили операцию сходства в виде \cap):

$$\langle \quad \rangle \odot \langle \quad \rangle \quad \langle \cap \cap \cap \rangle$$

Во второй части работы будет показано, что построенный предикат выразим в Граф-АФП, тем самым мы покажем что Граф-АФП также допускает решение этой задачи.

В третьей части работы будет показано, как с использованием оператора int из Граф-АФП мы можем переходить от объекта g к его внутренней структуре C , и тем самым мы установим связь между ЛО и ДСМ-методом, что позволит естественным образом задавать более тонкие правила для предиката сходства с композицией отношений.

1. О построении предиката сходства с композицией отношений в ДСМ-методе

1.1. Предварительные обозначения, понятия и определения

Опишем алфавит.

- Константы: (1) Константные символы для объектов (без штрихов) и подобъектов (со штрихами): $C, C', C_1, C_1', C_2, C_2', \dots$; (2) Константные символы для признаков: D, D_1, D_2, \dots ; (3) Константные символы для бинарных отношений: R, S, R_1, S_1, \dots ;

- Переменные: (1) Переменные для объектов и подобъектов: V, V', X, X', Z, Z' (быть может, с нижними индексами); (2) Переменные для признаков: U, W, Y (быть может, с нижними индексами); (3) Переменные для бинарных отношений: r, s (быть может, с нижними индексами);
- Логические операторы Россера-Тюркетта: $\neg, \vee, \wedge, \rightarrow, \leftrightarrow$ и \forall, \exists , где $v \in \{+1, -1, 0\}, n \in \mathbb{N}$;
- Логические связки (внешние): Λ, \cup, \cap ;
- Вспомогательные символы: $(,), ,, \dots$.

Обозначение.

$\langle \rangle$

где Φ – формула.

В выражении $C \Rightarrow_1 D$, которое является внутренней формулой (т.е. принимает истинностные значения $\langle v, n \rangle$ и $\langle t, n \rangle$, где $v \in \{+1, -1, 0\}, n \in \mathbb{N}$), говорится, что объект C обладает множеством свойств D . В выражении $C \Rightarrow_2 D$, которое является внутренней формулой, говорится, что подобъект C есть причина (каузально вынуждает) множества свойств D .

В выражении $C_1 R C_2$, которое является внешней формулой (т.е. принимает классические истинностные значения t и f), говорится, что объекты C_1 и C_2 находятся в отношении R , т.е. $\langle C_1, C_2 \rangle \in R$.

Определим три универсума:

- $U^{(1)}$ – множество элементов, из которых образуются объекты и подобъекты;
- $U^{(2)}$ – множество заданных свойств, присущих объектам;
- $U^{(rel)}$ – множество бинарных отношений между объектами. Всех их мы считаем антирефлексивными и асимметричными.

1.2. Предикат сходства с композицией отношений

1.2.1. Модельная задача. Опишем модельную задачу, в которой будет учитываться композиция между отношениями, а именно будут изучаться множество людей, для каждого из которых хотя бы в одной родословной ветви сохраняется изучаемая болезнь. Для данной задачи ниже будет построен соответствующий предикат сходства.

Пусть у нас есть множество G , состоящее из непересекающихся множеств:

$$U \cup U'$$

где G_1 множество мужчин и женщин, G_2 множество детей людей из G_1 , G_3 множество детей людей из G_2 . Также даны два отношения R_1 и R_2 , а именно, выражение $C_1 R C_2$ означает, что человек $C_1 \in G_1$ есть родитель

человека $C_2 \in G_2$, а выражение $C_2 RC_3$ означает, что человек $C_2 \in G_2$ есть родитель человека $C_3 \in G_3$. Сами люди C могут быть выражены, например, в виде ДНК, т.е. посредством трёхмерных помеченных графов.

Пусть мы построили положительный предикат сходства с композицией отношений \mathcal{R} . Тогда выполнение предиката означает, что нашлись ≥ 2 человека из G_3 , имеющих общую часть ДНК C и общие симптомы D , для каждого из которых найдётся хотя бы один родитель из G_2 , причём количество всех найденных родителей ≥ 2 , и каждый из них также имеет общую часть ДНК C и общие симптомы D , и у каждого найденного родителя из G_2 также найдётся хотя бы один родитель из множества G_1 , причём количество всех найденных родителей ≥ 2 , и каждый из них также имеет общую часть ДНК C и общие симптомы D .

1.2.2. Положительный предикат сходства с отношениями. Положительный предикат сходства с композицией отношений \mathcal{R} будет строиться на основе положительного предиката сходства с отношениями \mathcal{R} , который в свою очередь будет строиться на основе предиката сходства без отношений \mathcal{R} . Для наглядности мы выберем конкретный предикат сходства без отношений, а именно обратный положительный предикат сходства \mathcal{R}^{-1} , отрицательная версия которого была представлена в [Финн, 2009] (перепечатана из сборника 1999 года). Такой выбор обусловлен тем, что ниже мы будем обобщать результат выразимости предиката сходства в Граф-АФП из работы [Кузнецов, 2006], в которой был использован именно он. Конечно, идеи, которые будут приведены ниже, будут справедливы и для остальных предикатов сходства, а также для их усилений, например, для запрета на контрпримеры (ниже мы также дадим это усиление). Так, если мы хотим подчеркнуть, что идея применима для всех предикатов сходства, то мы будем писать просто \mathcal{R} .

Предварительно сделаем пояснения относительно индексов у константных символов, а именно, константа « $C^{t,j}$ » мыслится стоящей в области определения (индекс «1») отношения \mathcal{R} (индекс « j »), а константа « $C^{2,j}$ » – в области значения (индекс «2») отношения \mathcal{R}_j (индекс « j »). Аналогичные соображения верны для констант D и переменных X и Y .

Опишем части предиката сходства с отношениями \mathcal{R} .

Часть

Λ

а также симметричная ей \sim заменяет условие рассмотрения кортежей объектов $\langle C_1, C_2 \rangle$, которые являются «объектами» для предиката сходства \sim , а операцией сходства \odot является поэлементное сходство объектов, т.е.:

$$\langle \quad \rangle \langle \quad \rangle \langle \quad \cap \quad \cap \quad \rangle$$

Вместо этого мы по-отдельности для множеств первых и вторых компонент строим предикат сходства, после чего соединяем их вместе, в частности, этим условием.

Теперь опишем, как нужно изменить $(\exists\exists)^+$ из предиката \sim , если объекты $\langle C_1, C_2 \rangle$ берутся из области определения отношения R_j . Т.к. теперь на объекты накладывается дополнительное условие в виде \sim , имеющий вид $\langle C_1, C_2 \rangle \in R_j$, то, представив $(\exists\exists)^+$ в виде $\langle C_1, C_2 \rangle \in R_j \wedge \langle C_1, C_2 \rangle \in \sim$, мы видим, что его необходимо заменить на:

$$\wedge$$

получив $(\exists\exists)^{1,+}$ (конечно, т.к. само Ψ^+ имеет вид импликации, то, используя эквиваленции классической логики, мы можем поместить описанную добавочную формулу в посылку импликации Ψ^+). Симметрично добавим

$$\wedge$$

в $(\exists\exists)^+$, получив $(\exists\exists)^{2,+} \Psi^{2,+}$.

Примечание. Представленная здесь идея верна и для усилений предикатов сходства. Например, для запрета на контрпримеры, который имеет вид $\langle C_1, C_2 \rangle \in R_j \wedge \langle C_1, C_2 \rangle \in \sim$, мы получаем (она добавляется к \sim (см. ниже) через \wedge):

$$\wedge$$

Теперь определим предикат \sim как предикат \sim без кванторной приставки (т.к. в дальнейшем мы будем иметь дело с обратным предикатом сходства, то кванторная приставка будет иметь вид « \forall », с дополнительным условием $\langle C_1, C_2 \rangle \in R_j$ и $\Psi^{1,+}$ вместо Ψ^+ . Симметрично определим \sim .

Т.к. ранее мы договорились рассматривать «объекты» $\langle C_1, C_2 \rangle$ покомпонентно, то теперь, помимо объединения посредством \vee и \wedge , их необходимо соединить кванторно (в \forall (симметрично для \exists)), помимо переменных до R_j , которые соответствуют переменным для R_i , за счёт появились переменные после R_j).

$$\begin{array}{c} \wedge \\ \wedge \end{array}$$

Теперь определим \sim как

\wedge

На основе положительного предиката \sim может быть построен отрицательный предикат путём замены +1 на -1 в $\langle \sim \rangle$ (или могут браться другие комбинации предикатов сходства с их усилениями, например, как предложено в работе [Финн, 2009], взять простой положительный и обратный отрицательный), что позволит породить п.п.в. I-го рода с отношениями (для наглядности выпишем только один):

$$\begin{array}{c} \Rightarrow \qquad \qquad \qquad \wedge \\ \hline \langle \sim \rangle \qquad \Rightarrow \end{array}$$

Аналогичные соображения для п.п.в. I-го рода верны и для предиката сходства с композицией отношений.

1.2.3. Положительный предикат сходства с композицией отношений. Теперь перейдём к описанию частей \sim .

Вместо части

$$\begin{array}{c} \wedge \qquad \qquad \wedge \end{array}$$

в которой говорится, что искомые объекты $\langle C_1, C_2 \rangle$ из области значения отношения R_1 и объекты $\langle C_1, C_2 \rangle$ из области определения отношения R_2 , должны совпадать, мы выберем верхнюю кванторную приставку (можно было и нижнюю).

В части

\wedge \wedge

\wedge \wedge

схватывается идея родословной ветви. Заметим, что в данном условии уже содержится условие .

Распишем, как нужно изменить $(\exists Z)^+$ для каждой из области G:

(1) Для области G_1 необходимо добавить

\wedge

чтобы получить нужный вид ограничения .

(2) Для области G_2 необходимо добавить как

\wedge

чтобы получить нужный вид ограничения , так и

\wedge

чтобы получить нужный вид ограничения .

(3) Для области G_3 необходимо добавить

\wedge \wedge \wedge

чтобы получить нужный вид ограничения . Заметим, что при добавлении мы также получаем вид ограничения .

Формулы , и определяются как выше; отметим только, что в нужно добавить как так и , а в достаточно добавить (т.к. автоматически будет выполнено).

В части

 \wedge \wedge \wedge \wedge

говорится об устойчивости паттерна $C' \Rightarrow_2 D$ при наследовании.

Теперь определим

как

2. О выразимости предиката сходства с композицией отношений в Граф-АФП

2.1. Предварительные обозначения, понятия и определения

Анализ формальных понятий (АФП) – ветвь алгебраической теории решёток, метод анализа данных и модель машинного обучения, основанного на отношениях поглощения (порядка общности) [Kuznetsov, 2019], [Ganter et al., 1999]. Одним из расширений АФП является Граф-АФП [Ferré et al., 2020].

Определение 2.1. Графовым контекстом в Граф-АФП называют тройку $K=(G,M,I)$, где G – множество объектов, $M=\{M^1, M^2, \dots, M^n\}$ – множество признаков, разбитых на подмножества M^i , $I \subseteq G^* \times M$ – отношение инцидентности между кортежами объектов и признаками, где i местный кортеж может иметь признак только из множества M^i . \mathcal{M}^i – множество переменных, \subseteq .

Определение 2.2. Спроектированный графовый узор (СГУ) есть пара (G, π) , где G — графовый узор (ГУ), а π — кортеж переменных, который мы будем называть **проецирующим кортежем**. π есть **местность** СГУ.

Мы будем **обозначать** множество СГУ через \mathcal{G} , а через \mathcal{G}_k подмножество k-СГУ, т.е. СГУ, имеющих местность k.

Определение 2.3. k местным отношением объектов R , т.е. $|R|=k$, будем называть подмножество k местных кортежей, т.е. $R \subseteq G^k$.

Мы будем **обозначать** через \mathcal{M} множество всевозможных отношений, а через \mathcal{M}^k подмножество k местных отношений.

Для произвольных M^1, M^2 и I определим операторы int и ext :

$$\begin{aligned} \cap \\ \subseteq \end{aligned}$$

где СГУ \mathcal{M} есть описание кортежа объектов \mathcal{M} .

В дальнейшем мы ограничимся случаем графового контекста с $M=\{M^1, M^2\}$. Договоримся вместо операторов ext и int писать просто \cap , если $K=(G, \{M^1, M^2\}, I)$ – графовый контекст. Также введём два оператора:

- Оператор \cap , действие которого на СГС определим как:
- Оператор \subseteq , действие которого на $A=\{w, w_n\} \subseteq M^k$ определим как:

где \mathcal{M} .

Утверждение 2.1. \cap \cap .

Утверждение 2.2. \subseteq \subseteq .

Договоримся, если не оговорено противное, считать, что если мы пишем \mathcal{M} , где \mathcal{M} и \mathcal{M} , то \mathcal{M} .

Предварительно переведем множество бинарных отношений $U^{(\text{rel})}$ в множество $U^{*(\text{rel})}$, элементами которого будут такие объекты R^* , что $R^* = \{g^1, g^2\}$, где $g^1 \subseteq g^2 \times g^2$.

Примем предположение о том, что примеры являются либо положительными, либо отрицательными относительно каждого функционального (целевого) признака, т.е. элемента из $U^{(2)}$. Эту ситуацию можно представить следующим образом:

- Графовый контекст структуры и отношений $K_M(G, \{M^1, M^2\}, I)$: (1) M^1 есть множество структурных признаков, $M^1 = U^{(1)}$, т.е. для объекта $g \in G$ множество $M^1(g)$ есть множество элементов, из которых он состоит, т.е. $M^1(g) = \{g^1\}$ для некоторого $g^1 \in G$; (2) M^2 есть множество отношений, $M^2 = U^{*(\text{rel})}$, т.е. для упорядоченной пары $(g^1, g^2) \in G \times G$ множество $M^2(g^1, g^2)$ есть множество отношений, в которых состоит пара (g^1, g^2) , т.е. $M^2(g^1, g^2) = \{g^1, g^2\}$ для некоторых $g^1, g^2 \in G$, где $g^1 \subseteq g^2 \times g^2$ и $C^1 R_1 C^2, \dots, C^1 R_k C^2$.

- Формальный контекст свойств $K_P(G, P, J)$, где P есть множество свойств, $P = U^{(2)}$ и $J = \{+, -\}$, т.е. для объекта $g \in G$ объектное содержание $\{g\}^+$

есть множество свойств, которыми он обладает, а объектное содержание $\{g\}^-$ есть множество свойств, которыми он не обладает, т.е., если \Rightarrow , то $\{g\}^+=D$, а если \Rightarrow , то $\{g\}^-=D$ для некоторого .

Пусть A – множество, а P, Q – бинарные отношения. Тогда определим:

- $A \circ Q := \langle y \rangle \in A \wedge y$;
- $P \circ A := \langle y \rangle y \in A \wedge y$;
- $P \circ A \circ Q := \langle x, y \rangle z (z \in A \wedge x P z \wedge z Q y)$.

2.2. Теорема о выразимости обратного положительного предиката сходства с композицией отношений в Граф-АФП

Сформулируем теорему, которая будет являться обобщением теоремы из [Кузнецов, 2006]. Договоримся, что мы не различаем объект g и одно-местный кортеж (g) когда из контекста ясно, что должно быть.

Теорема 2.1 ((+)-обратный метод с композицией отношений через соответствия Галуа). Для произвольных $V \subseteq M^1$, $W \subseteq P$, $V \neq \emptyset$, $W \neq \emptyset$, следующие два утверждения эквивалентны:

Имеет место $|E_i| \geq 2$ и \subseteq и \subseteq , где $i \in \{1, 2, 3\}$, а также \subseteq и \subseteq и \subseteq и \subseteq и \subseteq , где $A = \text{dom}(R_1 \circ E_2)$, $B = \text{rng}(E_1 \circ R_1)$, $C = \text{dom}(R_2 \circ E_3)$ и $D = \text{rng}(E_1 \circ (R_1 \circ E_2 \circ R_2))$, а E_i находятся из решения следующей теоретико-множественной системы уравнений:

1. $E_1 = W^+ \cap A$.
2. $E_2 = W^+ \cap B \cap$.
3. $E_3 = W^+ \cap D$.

Если мы добавим запрет на контрпримеры, то во вторую часть теоремы выше к выполняемым требованиям мы должны добавить \subseteq и \subseteq и \subseteq и \subseteq и \subseteq .

3. Об усилении предиката сходства с композицией отношений средствами ЛО

Логика описания (ЛО) (описательная логика, дескрипционная логика) является одним из языков представления знаний, позволяющая создавать онтологии [Rudolph, 2011], [Baader et al., 2017].

Сделаем важное **наблюдение**. В ДСМ-методе мы имеем дело с внутренней структурой S объекта g , однако средств ДСМ-метода недостаточно чтобы перейти к самому объекту g ; в ЛО мы имеем дело с объектом g , однако средств ЛО недостаточно чтобы перейти к его внутренней структуре S . АФП, через оператор int , позволяет переходить от g к S , и наоборот.

рот (в силу уникальности внутренних структур). Если внутренняя структура C имеет вид, отличный от множества, необходимо воспользоваться оператором int из [Ferre, 2023].

Запишем с использованием Граф-АФП и ЛО (т.к. интерпретация фиксирована, аксиомы ЛО принимают классические истинностные значения t и f). Предварительно необходимо произвести перевод отношений R из множеств упорядоченных пар вида $\langle C_1, C_2 \rangle$ в множество упорядоченных пар вида $\langle g_1, g_2 \rangle$.

$$\begin{array}{ccccccc}
 & \wedge & & \wedge \wedge & & \wedge & & \wedge \\
 \wedge & & \wedge & & \wedge \wedge & & \wedge & & \wedge \\
 & & & & & & & & \wedge
 \end{array}$$

Мы можем усилить аксиому ЛО, заменив « R_j » на, например, « $\geq 2R_j$ », что будет означать, что мы смотрим на тех людей из G , у которых как минимум два ребёнка унаследовали изучаемую болезнь.

Заключение

Результаты, полученные в настоящей работе имеют как локальный характер, а именно показано, что средствами ДСМ-метода и Граф-АФП можно порождать обоснованные гипотезы о причине в задаче наследственной болезни, так и глобальный. Последнее выражается, во-первых, в том, что мы построили предикат сходства с отношениями, как следствие, теперь средствами ДСМ-метода можно решать задачи в которых без учёта отношений мы бы получали неправильный результат. Во-вторых, мы обогатили Граф-АФП. В-третьих, мы решили задачу объединения ДСМ-метода и ЛО.

В дальнейшем автор планирует создать необходимую программу, в которой было бы реализовано всё то, что было описано в настоящей работе, и провести следующее сравнение: насколько лучше будет проявлять себя ДСМ-метод, если в модельной задаче мы будем учитывать наследственность, а именно (рис. 1), сравнить случай 1 (ДСМ-метод без отношений), случай 2 и случай 3 (ДСМ-метод с одним отношением), случай 4 (ДСМ-метод с двумя отношениями) и случай 5 (ДСМ-метод с композицией двух отношений).

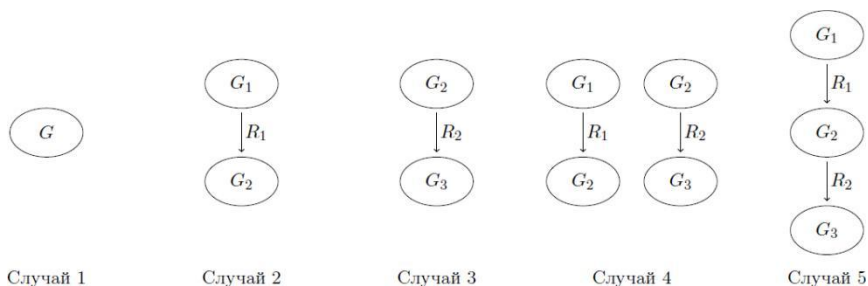


Рис. 1

Благодарности. Автор выражает благодарность д.т.н., профессору, В.К. Финну за указание, что построенные предикаты являются предикатами сходства; их «объектами» являются кортежи объектов, а их сходством является поэлементное сходство объектов.

Список литературы

- [Кузнецов, 2006] Кузнецов С.О. ДСМ-метод на языке соответствий Галуа // Научно-техническая информация. Серия 2: Информационные процессы и системы. – 2006. – №. 12. – С. 1-7.
- [Финн, 2009] Финн В.К. Синтез познавательных процедур и проблема индукции // Научно-техническая информация. Серия 2: Информационные процессы и системы. – 2009. – № 6. – С. 1-37.
- [Baader et al., 2017] Baader F. et al. Introduction to description logic. – Cambridge University Press, 2017.
- [Ferré et al., 2020] Ferré S., Cellier P. Graph-FCA: An extension of formal concept analysis to knowledge graphs // Discrete applied mathematics. – 2020. – Vol. 273. – P. 81-102.
- [Ferré, 2023] Ferré S. Graph-FCA Meets Pattern Structures // International Conference on Formal Concept Analysis. – Cham: Springer Nature Switzerland, 2023. – P. 33-48.
- [Ganter et al., 1999] Ganter B., Wille R. Formal Concept Analysis: Mathematical Foundations. – 1999.
- [Kuznetsov, 2019] Kuznetsov S.O. Ordered Sets for Data Analysis // arXiv preprint arXiv:1908.11341. – 2019.
- [Rudolph, 2011] Rudolph S. Foundations of description logics // Reasoning Web International Summer School. – Berlin, Heidelberg: Springer Berlin Heidelberg, 2011. – P. 76-136.