Федеральное государственное автономное образовательное учреждение
высшего образования
«Московский физико-технический институт
(национальный исследовательский университет)»
Физтех-школа Прикладной Математики и Информатики
Кафедра интеллектуальных систем

Направление подготовки / специальность: 03.03.01 Прикладные математика и физика (бакалавриат)
Направленность (профиль) подготовки: Компьютерные технологии и интеллектуальный анализ данных

ОПРЕДЕЛЕНИЕ МИНИМАЛЬНОЙ ДОПУСТИМОЙ ФУНКЦИОНАЛЬНОЙ КАТЕГОРИИ ПО НАБОРУ ПРЕЦЕДЕНТОВ

(бакалаврская работа)

Студент:
Кульков Александр

____
(подпись студента)

Научный руководитель:
Рудаков Константин Владимирович,
д-р физ.-мат. наук, акад. РАН

____
(подпись научного руководителя)

Консультант *(при наличии)***:**
____
____
(подпись консультанта)

Москва 2020

Содержание

Аннотация

В данной работе исследуется вопрос построения допустимых функциональных сигнатур, определяющих отображения, переводящие один заданный набор кортежей в другой. Проверка допустимости функциональной сигнатуры сводится к определению транзитивной замкнутости ориентированного графа, а также определению изоморфных подграфов в графах такого вида. Таким образом, задача нахождения сигнатуры с наименьшим числом используемых функций сводится к нахождению изоморфных друг другу подграфов в графе, построенном по заданным областям зависимости.

1 Введение

В середине 70-х годов академиком РАН Ю. И. Журавлёвым был предложен алгебраический подход к синтезу корректных алгоритмов. В дальнейшем его идеи получили своё развитие в работах К. В. Рудакова, предложившего алгебраическую теорию универсальных и локальных ограничений.

В рамках алгебраической теории общая задача распознавания образов сводится к построению специфических отображений, на которые дополнительно накладываются так называемые локальные (прецедентные) и универсальные ограничения. Среди рассматривемых типов универсальных ограничений выделяются функциональные ограничения, в рамках которых устанавливается требование к отображениям, чтобы они по форме удовлетворяли некоторой сигнатуре.

По определению, универсальные ограничения должны образовывать морфизмы некоторой категории. Критерии, позволяющие определить, выполнено ли это условие были ранее сформулированы К. В. Рудаковым в [7]. Основной целью данной работы является дальнейшее исследование конструкций, связанных с функциональными ограничениями. В частности, критерии того, что отображения, удовлетворяющие функциональным ограничениям, являются морфизмами некоторой категории, переформулируются в общих терминах теории графов. Данная переформулировка позволяет под новым углом взглянуть на задачу построения ограничений, согласующихся с некоторым заданным набором прецедентов.

Основными результатами работы является демонстрация NP-полноты исследуемой задачи, а также предложенный эвристический алгоритм её решения. Насколько известно автору работы, постановка задачи нова и не исследовалась ранее. Работа разделена на два основных блока, первый из которых посвящён введению основных терминов и постановке задачи, а второй – непосредственному изложению полученных результатов.

2 Постановка задачи

2.1 Основные понятия и определения

Определение 2.1. *Пусть $f : \mathbb{Z} \mapsto U$ – некоторая функция, определённая на множестве целых чисел. Будем сокращённо обозначать индексированное упорядочен-*

ное мультимножество $(f_l, f_{l+1}, \ldots, f_r)$ *как* $(f_k)_{k=l}^{r}$. *Составленное аналогичным образом неупорядоченное мультимножество* $\{f_l, f_{l+1}, \ldots, f_r\}$ *будем обозначать как* $\{f_k\}_{k=1}^{r}$.

Определение 2.2. *Через* $S(X)$ *будем обозначать группу всех перестановок множества* X. *В частности, для* $X = \{k\}_{k=1}^{n}$ *будем использовать обозначение* S_n, *а само множество* $\{k\}_{k=1}^{n}$ *будем обозначать как* X_n.

Определение 2.3. ***Функциональной сигнатурой*** $\varphi = (A_1, \ldots, A_n, \lambda)$ *называется семейство* $(A_k)_{k=1}^{n}$ *упорядоченных подмножеств* X_n *вместе с функцией* λ:

$$\lambda : X_n \mapsto X_t, \tag{2.1}$$

где $t \leqslant n$. *При этом для любых* i *и* j *из* X_n *должно быть выполнено:*

$$\lambda_i = \lambda_j \implies |A_i| = |A_j| \tag{2.2}$$

Мощность множества A_i будем обозначать как $z(\lambda_i) = |A_i|$, а его элементы – как набор $A_i = (\xi_{i,k})_{k=1}^{z(\lambda_i)}$. Введённые таким образом наборы A_i представляют собой описания «областей зависимости» элементов под действием отображения, описываемого сигнатурой. Функция λ, в свою очередь, задаёт для каждого элемента конкретный номер функции, которая будет применяться, чтобы вычислить значение данного элемента. Таким образом, ограничение (2.2) задаёт условие на то, что если некоторые два элемента порождаются одной и той же функцией, то количество аргументов, передаваемых данной функции, должно совпадать. Формально отображения, определяемые заданной функциональной сигнатурой можно определить так:

Определение 2.4. *Отображение* $f : U^n \mapsto V^n$ *называется* φ*-отображением если существует набор* $(f_k)_{k=1}^{t}$, *такой что* $f_k : U^{z(k)} \mapsto V$ *и* $f((u_i)_{i=1}^{n}) = (v_j)_{j=1}^{n}$, *где:*

$$v_j = f_{\lambda_j}\left((u_{\xi_{j,i}})_{i=1}^{z(\lambda_j)}\right) \tag{2.3}$$

Определение 2.5. **Категорией** *называется класс* **объектов**, *таких что:*

1. *Для любой пары объектов* A *и* B *задано множество* **морфизмов** $\hom(A, B)$;

2. *Для любой пары* $f \in \hom(A, B)$ *и* $g \in \hom(B, C)$ *определена их* **композиция**:

$$g \circ f \in \hom(A, C), \tag{2.4}$$

$$(a \circ b) \circ c = a \circ (b \circ c) \tag{2.5}$$

3. Для любого A есть тождественный морфизм $id_A \in \hom(A, A)$, такой что:

$$\forall f \in \hom(A, A) : f \circ id_A = id_A \circ f = f \tag{2.6}$$

Как было сказано ранее, ограничения, которые накладываются функциональными сигнатурами, должны играть роль функциональных ограничений в алгебраической теории, поэтому интерес представляют только такие сигнатуры, что порождаемые ими отображения можно интерпретировать как множество морфизмов некоторой категории. То есть, данная сигнатура должна допускать тождественные отображения, а также композиция отображений, удовлетворяющих сигнатуре, также должна ей удовлетворять. Такие сигнатуры называются *допустимыми*.

2.2 Симметрические сигнатуры

Так как порядок аргументов для функций является скорее некоторой условностью, имеет смысл также рассмотреть сигнатуры, которые не задают жёстких ограничений на порядок элементов в областях зависимости.

Определение 2.6. Симметрической сигнатурой $\theta = (D_1, \dots, D_n, \mu)$ *называется совокупность $(D_k)_{k=1}^n$ неупорядоченных подмножеств X_n вместе с функцией*

$$\mu : X_n \mapsto X_t, \tag{2.7}$$

где $t \leqslant n$. При этом для любых i и j из X_n должно быть выполнено:

$$\mu_i = \mu_j \implies |D_i| = |D_j| \tag{2.8}$$

Как и в случае обычных функциональных сигнатур мы будем использовать обозначение $z(\mu_i) = |D_i|$ для размера областей зависимости, а также $D_i = \{\eta_{i,k}\}_{k=1}^{z(\mu_i)}$ для обозначения элементов этих множеств. Заметим, что, в силу неупорядоченности множеств, порядок, определяемый η_i является в некотором смысле произвольным и определяется лишь внутренним устройством структуры, которая используется для хранения множеств D_i.

Определение 2.7. *Отображение $f : U^n \mapsto V^n$ называется θ-отображением если существует набор перестановок $(\pi_j)_{j=1}^n$ и функций $(f_k)_{k=1}^t$, таких что $\pi_k \in S_{z(\mu_k)}$, $f_k : U^{z(k)} \mapsto V$ и $f((u_i)_{i=1}^n) = (v_j)_{j=1}^n$, где*

$$v_j = f_{\mu_j}\left(\left(u_{\eta_{j,\pi_j(i)}}\right)_{i=1}^{z(\mu_j)}\right) \tag{2.9}$$

Таким образом, перестановка π_j задаёт порядок, в котором аргументы должны будут передаваться отображению f_{μ_j}. Такой подход снимает необходимость определения порядка аргументов в сигнатуре и позволяет оперировать неупорядоченными множествами. Альтернативный подход к использованию симметрических сигнатур предоставляют симметрические функции.

Определение 2.8. *Отображение $f : U^n \mapsto V^n$ называется симметрическим θ-отображением если существует набор симметрических функций $(f_k)_{k=1}^t$, таких что $f((u_i)_{i=1}^n) = (v_j)_{j=1}^n$, где*

$$v_j = f_{\mu_j}\left((u_{\eta_{j,i}})_{i=1}^{z(\mu_j)}\right) \tag{2.10}$$

Тот факт, что значения симметрических функций не зависят от порядка, в котором им подаются аргументы, позволяет нам интерпретировать подаваемые им аргументы как множества и использовать в определении «служебную» нумерацию $(\eta_{j,k})_{k=1}^{z(\mu(j))}$. К сожалению, полная потеря информации о порядке аргументов, которая происходит в случае использования симметрических функций, не позволяет построить содержательный тождественный морфизм в большинстве случаев, поэтому такие отображения, как правило, не образуют категорию.

2.3 Формальная постановка

Задача 2.1. *Есть два набора из m кортежей по n элементов $(U_k)_{k=1}^m$ и $(V_k)_{k=1}^m$, где $U_k, V_k \in X_n$. Нужно найти допустимую (симметричную) функциональную сигнатуру θ, такую что существует θ-отображение, переводящее каждый кортеж U_k в соответствующий ему кортеж V_k. При этом из всех таких отображений нас интересуют такие, которые в первую очередь минимизируют суммарный размер областей зависимости, а во вторую – число t используемых в сигнатуре функций.*

3 Поиск допустимых сигнатур

3.1 Критерий категориальности

К. В. Рудаковым был сформулирован следующий критерий:

Теорема 3.1. *Для того, чтобы множество φ-отображений при заданной функциональной сигнатуре φ образовывало множество морфизмов некоторой категории,* *необходимо и достаточно, чтобы выполнялись следующие условия:*

1. *Условия существования тождественного отображения:*

   (a) $i \in A_i$ *для всех* $i \in X_n$*;*

   (b) *Если* $\lambda_i = \lambda_j$*, то* $i = \xi_{i,k} \iff j = \xi_{j,k}$*;*

2. *Условия замкнутости относительно суперпозиции:*

   (a) *Если* $i \in A_j$*, то* $A_i \subset A_j$*;*

   (b) *Если* $\lambda_i = \lambda_j$*, то* $\lambda_{\xi_{i,k}} = \lambda_{\xi_{j,k}}$ *для любого* k*;*

   (c) *Если* $\lambda_i = \lambda_j$*, то* $\xi_{\xi_{i,k},k_1} = \xi_{i,k_2} \iff \xi_{\xi_{j,k},k_1} = \xi_{j,k_2}$*.*

Данный результат позволяет за полиномиальное время проверить, является ли функциональная сигнатура допустимой. Среди его преимуществ можно выделить ярко выраженную «локальность» – чтобы произвести проверку достаточно смотреть лишь на области зависимости отдельных элементов, а также элементов из их области зависимости, не вдаваясь детально в общий вид порождённых сигнатурами структур. Такой подход, однако, становится затруднительным при оценке симметрических сигнатур, в то время как именно они являются более естественными для прецедентной постановки задачи.

Определение 3.1. *Назовём* **сигнатурным графом** *заданной* φ*-сигнатуры ориентированный граф, устроенный следующим образом:*

1. *Граф содержит* n *вершин, соответствующих элементам кортежей;*

2. *Из вершины* u *есть ребро в вершину* v *если и только если* $v \in S_u$*;*

3. *Вершина* v *покрашена в цвет, определяемый значением* λ_v*.*

Аналогичным образом сигнатурные графы можно определить и для симметрических сигнатур. Ключевое различие между ними заключается в том, что в обычных сигнатурных графах порядок рёбер, исходящих из вершины зафиксирован (т. е., задаётся кортежем), а в симметрических графах – нет (т. е., задаётся множеством).

Определение 3.2. *Пусть* $G = \langle V, E\rangle$ *– некоторый граф и* $U \subset V$ *– подмножество его вершин. Подграф, вершинами которого являются вершины* U*, а рёбрами – рёбра исходного графа, концами которых являются вершины из* U*, называется* **индуцированным** *множеством* U*.*

Определение 3.3. *Пусть G – некоторый граф, v – его вершина, $R(v)$ – множество вершин, достижимых из v. Подграф, индкцированный множеством $R(v)$ будем называть* **подграфом достижимости** *вершины v.*

Далее нам понадобятся некоторые основные определения из теории графов, с которыми можно подробнее ознакомиться, например, в [6].

Определение 3.4. *Пусть G и G' – два графа. Они называются* **изоморфными** *если существует перестановка π вершин G, которая ставит их в соответствие вершинам G' таким образом, что сохраняются структурные свойства графа. В частности:*

1. $(u, v) \in E \iff (\pi_u, \pi_v) \in E'$, *где E и E' – рёбра графов;*
2. $c(u) = c(v) \iff c'(\pi_u) = c'(\pi_v)$, *где $c(\cdot)$ и $c'(\cdot)$ – цвета вершин в графах;*
3. *В случае несимметричных сигнатур – сохраняется порядок потомков вершины.*

Определение 3.5. **Транзитивным замыканием** *графа называется граф, построенный на том же множестве вершин, такой что ребро из u в v есть если и только если v достижима из u в исходном графе. В частности, из любой вершины должно быть проведено ребро (петля) в неё саму.*

Помимо этого в [1] было введено понятие транзитивного сокращения:

Определение 3.6. **Транзитивным сокращением** *графа называется граф, построенный на том же множестве вершин, такой что его транзитивное совпадает с транзитивным замыканием исходного графа. Из всех таких графов транзитивным сокращением называется граф с наименьшим числом рёбер.*

Определение 3.7. *Граф называется* **транзитивно замкнутым** *если он совпадает со своим транзитивным замыканием и* **транзитивно редуцированным** *если он совпадает со своим транзитивным сокращением.*

Утверждение 3.1. *Граф является транзитивно замкнутым если и только если:*

1. *Для любой вершины в графе есть петля из неё в нею саму;*
2. *Если в графе есть рёбра $a \to b$ и $b \to c$, то в нём также есть ребро $a \to c$.*

Утверждение 3.2. *Если в графе нет циклов, то его транзитивное сокращение единственно с точностью до изоморфизма.*

Очевидно, между сигнатурными графами и функциональными сигнатурами есть взаимно однозначное соответствие. Из этого следует, что критерии категориальности могут быть переформулированы на языке сигнатурных графов.

Теорема 3.2. *φ-отображения, задаваемые данным сигнатурным графом образуют множество морфизмов некоторой категории если и только если:*

1. *Граф транзитивно замкнут;*

2. *Подграфы достижимости одинаково окрашенных вершин изоморфны.*

Доказательство.

Докажем эквивалентность данных критериев исходным критериям категориальности функциональной сигнатуры. Условия 1.a и 2.a фактически эквивалентны транзитивной замкнутости сигнатурного графа. Действительно, в каждой вершине графа должна быть петля ($i \in A_i$), а условие $i \in A_j \implies A_i \subset A_j$ может быть переформулировано как $(c \in A_b) \wedge (b \in A_a) \implies c \in A_a$, что эквивалентно условиям транзитивной замкнутости.

Условия же 1.b, 2.b и 2.c соответствуют тому, что подграфы достижимости изоморфны. Действительно, условия 1.b, 2.b и 2.c (при их сужении на транзитивно замкнутые графы) эквивалентны тому, что требуемая для изоморфизма перестановка заключается в том, чтобы сопоставить друг другу вершины, стоящие на одинаковых позициях относительно вершин i и j. Другими словами, данная перестановка должна сопоставить $\xi_{i,k}$ и $\xi_{j,k}$. ∎

Критерий категориальности для симметрических графов выглядит так же, за исключением того, что порядок исходящих рёбер не имеет значения при рассмотрении вопроса об изоморфизме индуцированных подграфов.

3.2 Построение сигнатур

В процессе построения функциональных сигнатур можно выделить три основные подзадачи, которые необходимо успешно решить:

1. Выделение областей зависимости;

2. Выделение элементов, порождаемых одинаковыми функциями;

3. Выделение изоморфных подграфов.

Здесь мы предполагаем, что предоставленный набор прецедентов является достаточно обширным, чтобы из него можно было извлечь соответствующие функциональные зависимости. Для этого может использоваться достаточно разнообразный набор подходов, например, экспертная оценка, изучение линейных зависимостей с помощью метода главных компонент или метода Белсли, а также исследование статистической зависимости с помощью критерия хи-квадрат и других методов проверки гипотезы независимости. Указанные методы могут быть найдены, например, в [2], [3] и [4]. Стоит заметить, что так как мы предполагаем, что в каждой прецедентной паре используется одно и то же преобразование, мы не можем достоверно узнать зависимости, которые всплывают при, например, повторном применении того же преобразования. Однако, если задаться вопросом о нахождении наименьшего по включению семейства областей зависимости, которое согласуется с найденными взаимосвязями, мы придём к тому, что оно получается в точности взятием транзитивного замыкания от их графа зависимостей. Так как на сегодняшний день вопрос о наиболее эффективном построении исходных зависимостей остаётся открытым, мы не будем заострять на нём внимание и предоставим специалистам выбирать наиболее подходящий метод исходя из предпосылок конкретной задачи.

3.3 Выделение одинаковых функций

Следует отметить, что даже в рассматриваемом сейчас упрощении, при котором вопрос о построении оптимальных областей зависимости выносится за рамки решаемой задачи, она остаётся вычислительно трудоёмкой. В частности, можно показать, что даже для случая когда область зависимости каждого элемента содержит только его самого, задача выделения наименьшего числа функций, которые можно использовать для построения φ-отображения, согласующегося с прецедентами, является NP-полной. Рассмотрим формальную постановку данной подзадачи:

Задача 3.1. *Даны две $n \times m$-матрицы X, Y такие что $x_{ij} \in U^d$ и $y_{ij} \in V$. Нужно разбить строки матриц на t блоков, чтоб k-му блоку соответствовала f_k такая что $f_k(x_{ij}) = y_{ij}$ для всех i из блока. При этом нужно минимизировать число блоков.*

В таком виде результаты вычисления функций с одинаковым числом аргументов можно представить в виде двух матриц, строки которых соответствуют номеру элемента в преобразованном кортеже, а столбцы – номеру прецедента, в котором этот элемент был получен.

Теорема 3.3. *Задача 3.1 является NP-полной.*

Доказательство.

Принадлежность соответствующей задачи разрешимости классу NP очевидна. Покажем, что NP-полная задача о хроматическом числе графа сводится к данной. Пусть G – некоторый граф, вершины которого пронумерованы целыми числами от 1 до n. Рассмотрим частный случай задачи 3.1 с $d = 1$. Пусть $1 \leqslant i < j \leqslant n$. Рассмотрим матрицы X и Y такие что $x_{ij} = x_{ji} = i \cdot (n+1) + j$, при этом если между i и j есть ребро, то $y_{ij} = 0$ и $y_{ji} = 1$, а в противном случае $y_{ij} = y_{ji} = 1$. Таким образом, любые две строки будут обладать единственным общим аргументом, на котором они принимают одно и то же значение в том и только том случае, если между ними нет ребра в графе. То есть, две вершины могут оказаться в одном блоке если и только если они могут быть покрашены в один и тот же цвет в исходном графе. ■

Аналогичным образом можно данную задачу свести к задаче о хроматическом числе – для этого нужно провести рёбра между теми строчками, для которых найдётся аргумент, в котором они отличаются. Таким образом, оптимальным решением данной подзадачи будет использование алгоритмов, решающих задачу о раскраске.

3.4 Выделение изоморфных подграфов

Наконец, рассмотрим оставшуюся часть задачи, связанную с выделением изоморфных друг другу подграфов достижимости. В общем случае данная задача достаточно трудна – на сегодняшний день известно, что проверка изоморфности ориентированных ациклических графов общего вида эквивалентна проверке изоморфности произвольных графов, для которой не известно ни лежит ли она в классе P, ни является ли она NP-полной. Сформулированные ранее критерии для сигнатур общего вида позволяли успешно проверять такой изоморфизм, опираясь на то, что из вершины непосредственно проведены рёбра во все достижимые из неё вершины (то есть, на то, что граф транзитивно замкнут). В случае же симметрических сигнатур выполнить такую же процедуру не представляется возможным в связи с тем, что структура графа по умолчанию не подразумевает какой-либо «канонический» порядок на множестве исходящих из заданной вершины рёбер.

При этом по своей природе поставленная задача предполагает именно работу с симметрическими сигнатурами, так как прецеденты сами по себе не позволяют определить порядок, в котором функции должны подаваться аргументы, единственная реально извлекаемая из них информация – множество данных аргументов. В связи с этим предлагается эвристическая процедура, позволяющая приближённо проверять изоморфизм графов такого вида. Вместо того, чтобы сравнивать подграфы достижимости как ориентированные графы, можно сравнивать лишь общую схему вычислений, которая ими задаётся. Для этого можно «развернуть» данные подграфы в деревья, соответствующие всем возможным путям из исходной вершины, а затем производить сравнение полученных структур, а не исходных графов.

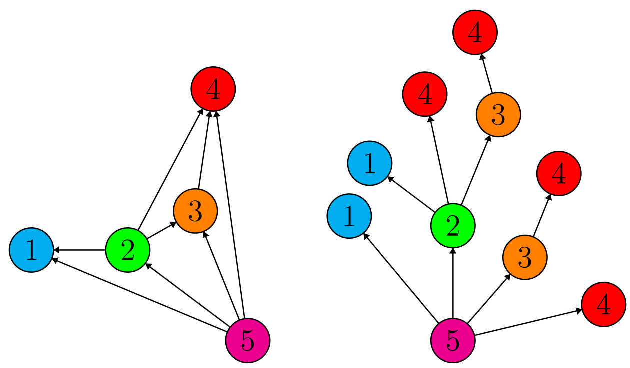

Рис. 1: Подграф достижимости и соответствующее ему дерево путей

Определение 3.8. **Деревом путей** *подграфа достижимости будем называть дерево, каждая вершина которого соответствует некоторому пути из корневой вершины подграфа достижимости и покрашена в цвет, соответствующий цвету вершины, в котором заканчивается этот пути.*

Определение 3.9. *Будем называть два подграфа достижимости* **псевдоизоморфными** *если изоморфны соответствующие им деревья путей.*

Критерии изоморфизма деревьев хорошо известны (см. [5], например):

Утверждение 3.3. *Два корневых ориентированных дерева изоморфны если и только если можно сопоставить непосредственных потомков корней первого и второго дерева друг другу таким образом, чтобы поддеревья сопоставленных вершин также были изоморфны.*

Для простоты ограничимся случаем когда сигнатурный граф ациклический. В случае циклических графов их компоненты сильной связности представляют собой клики, которые можно сжать и работать с конденсацией графа. При этом между кликами необходимо будет поддерживать дополнительное условие о том, что в любых двух кликах должно быть равное число вершин одного цвета. В таком случае можно рассматривать вершины в порядке увеличения размера соответствующих им деревьев путей, что будет гарантировать то, что вершины, которые можно покрасить в один и тот же цвет могут быть рассмотрены одновременно и при этом строго после того, как будут рассмотрены все достижимые из них вершины. Далее вводится требование, чтобы подграфы достижимости покрашенных в одинаковый цвет вершин были псевдоизоморфными. С учётом данного требования проверка двух вершин на то, что они псевдоизоморфны сводится к тому, что множества цветов вершин, в которые из них ведут рёбра, совпадают. Таким образом можно выделить следующий алгоритм для построения функциональной сигнатуры:

1. Любыми средствами определяются базовые зависимости между элементами;
2. От обнаруженных зависимостей берётся транзитивное замыкание;
3. Вершины рассматриваются в порядке увеличения размеров деревьев путей;
4. Для данного размера группируются вершины с изоморфными подграфами;
5. Для выделения в них одинаковых решается задача 3.1;
6. Процесс продолжается пока все вершины не будут окрашены.

4 Заключение

В работе была изучена задача построения минимальных сигнатур по заданным наборам прецедентов. В ходе работы критерии допустимости сигнатуры были переформулированы в терминах теории графов, введено понятие симметрических функциональных сигнатур. Также была показана NP-полнота задачи и предложена эвристическая процедура для её приближённого решения в случае когда сигнатурный граф является ациклическим. Среди основных направлений для дальнейшей работы можно выделить рассмотрение случая неациклических графов, а также возможность применения в данной задачи транзитивно редуцированных вместо транзитивно замкнутых, что потенциально позволило бы работать с графами существенно меньшего размера, что может привести к существенному уменьшению времени работы.

Список литературы

[1] Aho A. V. The transitive reduction of a directed graph / A. V. Aho, M. R. Garey, J. D. Ullman // SIAM Journal on Computing. — 1972. — Т. 1, № 2. — С. 131–137.
http:// https://doi.org/10.1137/0201008

[2] Belsley D. A. Regression Diagnostics / D. A. Belsley, E. Kuh, R. E. Welsch. — John Wiley & Sons, Inc., 1980. — jun.
http:// https://doi.org/10.1002%2F0471725153

[3] Best J. A guide to chi-squared testing. / J. Best, P. E. Greenwood, M. S. Nikulin // Biometrics. — 1998. — mar. — Т. 54, № 1. — С. 392.
http:// https://doi.org/10.2307%2F2534027

[4] Jolliffe I.T. Principal Component Analysis / I.T. Jolliffe. Springer Series in Statistics. — Springer New York, 2013.
http:// https://books.google.ru/books?id=-ongBwAAQBAJ

[5] Valiente G. Tree Isomorphism / G. Valiente // Algorithms on Trees and Graphs. — Berlin, Heidelberg: Springer Berlin Heidelberg, 2002. — С. 151–251.
http:// https://doi.org/10.1007/978-3-662-04921-1_4

[6] Ахо А. В. Структуры данных и алгоритмы: [пер. с англ.] / А. В. Ахо, Д. Э. Хопкрофт, Д. Д. Ульман. — Вильямс, 2010.
http:// https://books.google.ru/books?id=aD1rkgAACAAJ

[7] Рудаков К. В. Алгебраическая теория универсальных и локальных ограничений для алгоритмов распознавания: Дис. док. физ.-мат. наук / Вычислительный центр АН СССР. — 1992. — С. 274.
http:// http://www.ccas.ru/frc/thesis/RudakovDocDisser.pdf