М.И. Забежайло

О некоторых оценках сложности вычислений в ДСМ-методе. Часть II[1]

Аннотация. Обсуждаются возможности организации интеллектуального анализа данных (ИАД), обеспечивающего экспертизу различных типов представления исходных данных (множеств признаков, графов, числовых векторов) однородными средствами их обработки. ДСМ-метод рассматривается как платформа для организации ИАД при анализе различных типов данных.

Представлены некоторые оценки сложности вычислений, характерные для задач восстановления каузальных зависимостей из эмпирических данных. Показана наследуемость свойств трудно-разрешимости (NP-полноты и перечислительной полноты) ряда переборных задач при варьировании выразительных возможностей используемого языка представления исходных данных. Предложен детальный анализ оценок вычислительной сложности процедур ДСМ-анализа числовых данных, сформулированных в стиле метода сопутствующих изменений Д.С.Милля.

Ключевые слова: ДСМ-метод автоматического восстановления зависимостей из эмпирических данных, вычислительная сложность и оптимизация перебора, причинные зависимости в числовых данных, приближенный ДСМ-метод.

Введение

В первой части этой работы нами было начато обсуждение возможностей организовать интеллектуальный анализ данных (ИАД) так, чтобы он позволял обеспечивать обработку различных типов описания исходных данных (обычных или нумерованных множеств признаков, кортежей, графов, цепочек символов, числовых векторов) однородными средствами их обработки. Были рассмотрены некоторые процедурные особенности ДСМ-метода автоматического восстановления зависимостей из данных, позволяющие использовать его как платформу для организации ИАД.

Так же были представлены некоторые оценки сложности вычислений, характерные для задач восстановления каузальных зависимостей из эмпирических данных. Начато изучение наследуемости свойств трудно-разрешимости (NP-полноты и перечислительной полноты) ряда переборных задач при варьировании выразительных возможностей используемого языка представления исходных данных. Пришло время двинуться дальше.

Если взглянуть на обсуждаемую нами схему анализа данных в процессе обучения на примерах и контрпримерах как на одно из оснований для формирования платформы – *понятийной*, *методологической*, *дескриптивной* и *аргументационной*, *алгоритмической,* – позволяющей организовать интеллектуальный[2] анализ данных, то представляется вполне естественным проследить как именно уже рассмотренные выше на «простейшем[3]» примере (где используемый тип представления данных – множества признаков, а операция сходства – пересечение множеств) комбинаторные свойства проявляются при исполнении аналогичных (в том смысле, что при их анализе используется та же понятийная схема <*примеры* и *контрпримеры* => *формализация сходства* => *классы сходства* => *классы эквивалентности* => … >) операций и с другими типами данных.

Здесь мы следуем уже обсуждавшейся в [1] точке зрения на характерные особенности ИАД

[1] см. Забежайло М.И. О некоторых оценках сложности вычислений в ДСМ-методе // Искусственный интеллект и принятие решений. – 2015. – Часть I: № 1, С.3 - 17.

[2] В классическом смысле этого термина – [2] и др.

[3] Далее мы обратимся к операциям с более сложными типами данных.

(в смысле действительно *интеллектуального* анализа данных), которые позволяют выделять соответствующие компьютерные системы-«*усилители*» интеллектуальных возможностей эксперта (например, [3] и др.) из всего множества, в том числе - и «обычных»,- вычислительных инструментов компьютерного анализа данных. В обозначенном контексте представляется целесообразным обратить внимание, в частности, на следующие функциональные свойства используемых инструментов ИАД:

- возможности выбора адекватных средств представления данных об объекте анализа (т.е. возможности варьирования типов данных в рамках одного и того же класса алгоритмов интеллектуального анализа данных), например: множеств признаков, кортежей, графов, числовых векторов, …; а также
- возможности настройки локальных средств анализа эмпирических данных в изучаемой предметной области на работу с соответствующими типами данных (например, выбор соответствующей алгебраической операции сходства при формализации сходства анализируемых объектов – множеств признаков, кортежей, графов, числовых векторов, …).

Наличие (доступность исследователю, реализующему ИАД) таких средств позволяет

- формулировать классы однотипных формализованных моделей анализа данных, дающих возможности обеспечить варьирование инструментов принятия решений для их «настройки» на особенности данных в изучаемой предметной области (в том числе – варьирование как конкретных математических средств поддержки приятия решений, так и «подстройки» соответствующей алгоритмики);
- формировать (в частности – в процессе сравнения «наследуемости» порождаемых результатов, при переходе от более «простых» к более «сложным» способам представления данных) содержательные интерпретации результатов, получаемых применением используемых формализмов (в том числе – содержательной интерпретации порождаемых зависимостей – каузальных, ассоциативных и т.п.) а также оценки достаточности оснований для принятия полученных содержательных результатов (их обоснованности, корректной интерпретируемости и т.п.); и в случае неудовлетворительной ситуации в данном аспекте
- задействовать возможности анализа причин фиксируемой неадекватности, чтобы вернуться к исходным шагам проводимого интеллектуального анализа (выбору формальных средств представления данных и т.д. – см. выше) и провести, опираясь на выполненную «работу над ошибками», новую – более адекватную - «настройку» используемых решающих средств ИАД на критически важные особенности изучаемой предметной области.

Именно такой «цикл»:

- *обработка* данных,

- *оценка приемлемости* получаемых результатов и последующая

- *коррекция* (дополнительные «*настройки*») задействованного инструментария (математических моделей вместе с реализующими их алгоритмикой и программными системами),

дает *аргументированные* основания считать реализуемый подход действительно *интеллектуальным* анализом данных.

Теперь мы обратимся к обсуждению особенностей работы в рамках используемой схемы ИАД

исходные ***примеры*** и ***контрпримеры*** =>
=> формализация их ***представления*** (выбор средств описания данных) =>
=> формализация операции ***сходства*** =>
=> ***классы сходства*** =>
=> ***классы эквивалентности*** =>
=> ***прогноз свойств*** новых объектов

с другими типами данных:

- нумерованными множествами;
- кортежами множеств признаков;
- графами (множествами признаков с отношениями на таких множествах);
- цепочками (упорядоченными последовательностями символов);
- векторами числовых значений параметров.

5. Работа с нумерованными множествами, кортежами и графами[4]

Пусть $\mathbf{U}=\{a_1,a_2, \ldots ,a_n\}$ есть как и ранее алфавит образующих (при этом, не теряя общности, ранее мы считали, что в **U** нет одинако-

[4] Нумерация разделов и утверждений продолжает уже начатую в первой части настоящей работы нумерацию.

вых элементов), а $\mathbf{\Omega}=\{A_1,A_2,...,A_m\} \subseteq 2^{\mathbf{U}}\backslash\varnothing$ - множество объектов (т.е. непустых множеств образующих), построенных над **U** (т.е. множество объектов, построенных из образующих $a_1, a_2, \ldots, a_n$.). Теперь будем предполагать, что среди элементов множества **U** могут встречаться графически совпадающие (по аналогии с тем, что в множестве **Ω** также могут быть графически совпадающие элементы, имеющие тем не менее различные имена). Рассмотрим множество

$$\mathbf{U^*} = \{a^*_1, a^*_2, \ldots, a^*_{n^*}\} = \\ =\{<a_1,k_1>,<a_2,k_2>, \ldots ,<a_n,k_n>\},$$

построенное из множества **U** по следующим правилам:

- для каждого $a^*_i = a_i$ выделяются все его вхождения в универсум **U***,
- число таких вхождений k_i вместе с самой образующей a_i будет порождать каждую соответствующую пару - элементов множества **U***.

Определение 7

1. Множества вида $\mathbf{U} = \{<a_1,k_1>,<a_2,k_2>, \ldots ,<a_n,k_n>\}$, где каждая из входящих в него образующих характеризуется еще и числом (или частотой) таких вхождений, мы будем называть ***нумерованными множествами***.

2. Пусть заданы нумерованные множества $\mathbf{U_1} = \{<a_{11},k_{11}>, <a_{12},k_{12}>, \ldots, <a_{n1},k_{n1}>\}$ и $\mathbf{U_2} = \{<a_{21},k_{21}>,<a_{22},k_{22}>, \ldots ,<a_{n2},k_{n2}>\}$.

Их ***объединением*** назовем множество $\mathbf{U_{1\cup2}}$, в которое входят все образующие множеств $\mathbf{U_1}$ и $\mathbf{U_2}$, при этом частота вхождения каждой соответствующей образующей в объединение есть максимум из частот вхождения этой образующей в компоненты объединения (дополнительно считается, что не встречающаяся в нумерованном множестве образующая имеет нулевую частоту вхождения в него.):

$$\mathbf{U_{1\cup2}} = \{<a_i,k_i> \mid (<a_{1i},k_{1i}>\in\mathbf{U_1}) \,\&\, (<a_{2i},k_{2i}>\in\mathbf{U_2}) \\ \&\, (k_i = \boldsymbol{max}(k_{1i},k_{2i}) \,\&\, (1 \le \mathbf{i} \le \mathbf{n_1}+\mathbf{n_2})\}$$

Их ***пересечением*** назовем множество $\mathbf{U_{1\cap2}}$, в которое входят все общие образующие множеств $\mathbf{U_1}$ и $\mathbf{U_2}$, при этом частота вхождения каждой соответствующей образующей в пересечение есть минимум из частот вхождения этой образующей в компоненты объединения:

$$\mathbf{U_{1\cap2}} = \{<a_i,k_i> \mid (<a_{1i},k_{1i}>\in\mathbf{U_1}) \,\&\, (<a_{2i},k_{2i}>\in\mathbf{U_2}) \\ \&\, (k_i = \boldsymbol{min}(k_{1i},k_{2i}) \,\&\, (1 \le \mathbf{i} \le \mathbf{n_1}+\mathbf{n_2})\}$$

Вложимость нумерованного множества $\mathbf{U_1}$ в нумерованное множество $\mathbf{U_2}$ понимается естественным образом, т.е. $\mathbf{U_1} \subseteq \mathbf{U_2}$ означает, что $\mathbf{U_1} = \mathbf{U_{1\cap2}} = \mathbf{U_1} \cap \mathbf{U_2}$.

Учитывая идемпотентность, коммутативность и ассоциативность операций взятия максимума и минимума из пары натуральных чисел, а также известные алгебраические свойства обычных булевских операций, нетрудно убедиться, что справедлива

Лемма 11

Заданные **Определением 7** операции объединения и пересечения нумерованных множеств определяют, соответственно, верхнюю полурешетку с единицей (роль которой выполняет множество образующих алфавита **U** с максимальными значениями частот встречаемости соответствующих образующих) и нижнюю полурешетку с нулем (роль которого выполняет пустое множество) на соответствующем универсуме нумерованных множеств.

o

Определение 8

Нумерованное множество $A_{i1,i2,..,is} = A_{i1} \cap A_{i2} \cap \ldots \cap A_{is}$, где $\cap$ - операция пересечения нумерованных множеств, мы будем называть ***исчерпывающим сходством*** (пересечением) объектов из нумерованного множества **Ω** тогда и только тогда, когда:

(1) $A_{i1,i2,..,is} \neq \varnothing$, и

(2) $(\forall\mathbf{j})\ ([(A_j \in \mathbf{\Omega}) \,\&\, \neg(A_j \in \{A_{i1}, A_{i2}, \ldots, A_{is}\})] \supset [(A_j \cap A_{i1,i2,..,is}) \subset A_{i1,i2,..,is}])$[5]

o

Пусть заданы алфавит **U** и множество объектов (нумерованных множеств) над ним - **Ω**. По аналогии с **Определением 5** для данного случая также можно сформировать 10 соответствующих переборных задач.

Прежде чем обратиться к их детальному анализу рассмотрим возможность сводимости случая обработки нумерованных множеств к случаю обработки обычных множеств. Процедуру ***REFORM***$(\mathbf{Z_{num}},\mathbf{Z}(\mathbf{Z_{num}}))$ такого сведения задачи $\mathbf{Z_{num}}$ обработки нумерованных множеств

[5] Другими словами, каждое исчерпывающее пересечение $A_{i1,i2,..,is}$ - не пусто и образовано в точности всеми теми объектами из **Ω**, которые при пересечении дают именно множество $A_{i1,i2,..,is}$.

к порождаемой по ней (и аналогичной по формулировке) задаче $\mathbf{Z}(\mathbf{Z}_{num})$ обработки обыкновенных множеств задают следующие условия:

Процедура *REFORM*($\mathbf{Z}_{num}$, $\mathbf{Z}(\mathbf{Z}_{num})$)

Исходные данные:

- нумерованные множества $\mathbf{U}_{num}$ и $\mathbf{\Omega}_{num}$, условие задачи $\mathbf{Z}_{num}$. (Не теряя общности, мы будем считать, что универсум **U** есть обычное множество, т.е. образован парами вида $<a_i,1>$).

Алгоритм сведения:

- в исходном состоянии положим множества **U** и **Ω** пустыми;
- для каждой образующей a_i из $\mathbf{U}_{num}$ найдем в множестве $\mathbf{\Omega}_{num}$ объект с максимальным значением $k_i = k_{i0}$ - частоты вхождений образующей a_i в элементы $\mathbf{\Omega}_{num}$; затем
- внесем в множество **U** элементы $a^1{}_i, a^2{}_i$, ... , $a^{ki0}{}_i$ (ровно k_{i0} штук - по числу возможных ненулевых частот вхождений образующей a_i в пересечения элементов из $\mathbf{\Omega}_{num}$); далее
- для каждого элемента X_{num} из множества $\mathbf{\Omega}_{num}$ построим (заменой в X_{num} каждого элемента вида $<a_i,k_i>$ на множество элементов вида $a^1{}_i, a^2{}_i$, ... , $a^{ki}{}_i$) новый объект X, затем включим каждый построенный таким образом X в множество **Ω**. Построение описания задачи **Z** - тройки <**U**,**Ω**, *условие задачи* $\mathbf{Z}_{num}$> - завершено.

Взаимоотношения задач **Z** и $\mathbf{Z}_{num}$ описывает

Теорема 7

Любому исчерпывающему пересечению объектов из множества $\mathbf{\Omega}_{num}$ над универсумом $\mathbf{U}_{num}$ может быть сопоставлено одно и только одно исчерпывающее пересечение объектов из множества **Ω** над универсумом **U**.

Доказательство.

Пусть ω_{num} - исчерпывающее пересечение объектов из множества $\mathbf{\Omega}_{num}$ над универсумом $\mathbf{U}_{num}$. Построим множество ω, порождаемое из множество ω_{num} применением процедуры сведения ***REFORM*** (т.е. “перекодируем” каждое из кратных вхождений каждой из образующих из универсума $\mathbf{U}_{num}$ соответствующими однократными вхождениями образующих из универсума **U**). Легко видеть, что при такой “перекодировке” каждое исчерпывающее пересечение объектов из множества $\mathbf{\Omega}_{num}$ может быть сопоставлено его “образу” - исчерпывающему пересечению объектов из множества **Ω**. При этом построенное соответствие взаимно однозначно.

оо

Таким образом, **Теорема 7** позволяет нам свести вопросы о комбинаторных свойствах обсуждаемых переборных задач к воспроизведению уже разобранных нами свойств соответствующих комбинаторных проблем, возникших при обработке данных типа ***множества***. Другими словами, справедливо:

Следствие 13

Для нумерованных множеств остаются в силе все справедливые для работы с обычными множествами результаты, перечисленные выше **Теоремами 1-6** (вместе с соответствующими **Леммами** и **Следствиями**).

о

Итак, свойства трудно-разрешимости соответствующих комбинаторных задач наследуется при переходе к обработке данных, представленных нумерованными множествами.

Абсолютно аналогичными рассуждениями формируется утверждение о наследуемости (полученных для обработки данных, которые представлены множествами) результатов и на случай работы с кортежами (в том числе и нумерованных) множеств признаков. Работать с данными такого типа приходится, в частности, в случае анализа свойств физиологически активных веществ в ситуациях, когда химическая структура изучаемых соединений представлена, например, кодом ФКСП[6], а вместе с нею обрабатываются данные о тех или иных релевантных эффекту физико-химических характеристиках (например, концентрациях, растворимости, температурах и т.п.).

Здесь в каждом из «полей» соответствующего кортежа возникает ситуация уже рассмотренного нами выше типа (т.е. реализуется работа с обычными либо нумерованными множествами). Соответственно, на такие ситуации также распространяются (наследуются) свой-

[6] ФКСП (Фрагментарный Код Суперпозиции Подструктур) обеспечивает покрытие графа структуры изучаемого соединения соответствующими элементами множества стандартизованных подграфов, содержащего порядка 20 000 элементов (подробнее [4] и др.).

ства комбинаторной трудно-разрешимости соответствующих переборных задач.

В случае обработки данных, представленных цепочками (упорядоченными последовательностями, в том числе – с многократными вхождениями) символов из заданного исходного алфавита аналогичное (уже сформулированным выше в тексте) заключение о трудноразрешимости соответствующих комбинаторных проблем несложно получить, опираясь на следующее очевидное обстоятельство:

(**!**) «обычное» множество (символов из некоторого алфавита **U**) можно взаимно однозначным образом представить в виде упорядоченной последовательности символов (т.е. цепочки), перенумеровав образующие исходного алфавита **U** и выстроив в описании исходного множества его образующие в соответствии с упорядочением их номеров. (Таким образом, можно рассматривать «обычные» множества как подмножество всех таких непустых цепочек символов из **U**, которые не содержат повторных вхождений образующих).

При этом сходство цепочек определяется как операция выделения всех максимальных по вложению общих подцепочек для заданного множества цепочек.

Таким образом, справедливо

<u>Следствие 14</u>

Для цепочек символов конечного алфавита остаются в силе все справедливые для работы с обычными множествами результаты, определяющие вычислительную сложность задач

- ***КЛАСС ЭКВИВАЛЕНТНОСТИ РАЗМЕРА РОВНО k***
- ***ЧИСЛО КЛАСС ЭКВИВАЛЕНТНОСТИ.***

о

В случае, когда ИАД по обсуждаемой нами схеме необходимо вести с объектами (примерами и контрпримерами), представленными в виде графов, уже ожидаемые результаты о сложности задач

- порождения класса эквивалентности фиксированного размера $\boldsymbol{k} = \boldsymbol{k_0}$ и

- вычисления числа классов эквивалентности (порождаемых на заданном множестве исходных примеров и контрпримеров)

также могут быть получены на базе следующего простого соображения (о взаимной вложимости соответствующих классов переборных задач):

(**!!**) цепочки символов есть (специальный) частный вид (плоских) графов.

Таким образом, с учетом **Следствия 14** имеет место

<u>Следствие 15</u>

Для представления данных в виде графов остаются в силе все справедливые для работы с обычными множествами результаты, определяющие вычислительную сложность задач

- ***КЛАСС ЭКВИВАЛЕНТНОСТИ РАЗМЕРА РОВНО k***
- ***ЧИСЛО КЛАСС ЭКВИВАЛЕНТНОСТИ.***

о

6. Работа с числовыми векторами: базовые понятия и определения

Теперь обратимся к ситуации, когда ИАД ведется с объектами, представленными в виде векторов числовых значений соответствующих эмпирических параметров. Для уточнения операции сходства воспользуемся эвристикой Д.С. Милля, определяющей так называемый *метод сопутствующих изменений* [5]. В основу используемой формализации положена идея выделения участков (т.е. максимальных по вложению подмножеств столбцов) *ко-монотонного изменения* числовых значений множеств параметров (т.е. максимальных по вложению подмножеств строк) в исходно заданной матрице (подробнее [6, 7])

<**параметры** (строки)> x <**объекты** (столбцы)>,

где каждый анализируемый в рамках выполняемого ИАД объект представлен вектором числовых значений измеряемых характеристик (параметров). Здесь:

- в качестве (наделенных исследуемыми свойствами) *объектов* рассматриваются соответствующие *векторы численных значений параметров*;
- локально сходными объявляются каждые два объекта, для которых в рассматриваемом множестве *параметров* найдутся, по крайней мере, два таких, что их значения от одного из выбранных объектов к другому (по всему множеству включенных в рассмотрение объектов) *изменяются монотонно*;

- классы сходства (и покрывающие их классы эквивалентности) формируются всеми такими объектами, которые попадают в соответствующий "*участок монотонности* изменения значений параметров" в матрице всех исходно заданных для анализа объектов (т.е. в данном случае следует говорить о сходстве объектов в определенном <u>*контексте*</u> - множестве исходно заданных примеров. Именно такими *контекстными* зависимостями и будут обусловлены продемонстрированные далее отличия сложностных характеристик рассматриваемых задач ИАД для этого типа данных.).

Пример ИАД-обработки данных такого типа при анализе нефте-газоносных свойств территорий представлен в работе [8].

Итак, пусть **X** - матрица действительных чисел, образованная n столбцами и m строками:

$$\mathbf{X} = \mathbf{C} \times \mathbf{R} = \{C_1, C_2, ..., C_n\} \times \{R_1, R_2, ..., R_m\}.$$

Будем называть матрицу **X** универсумом, ее столбцы C_i (где $i \in \{1,2, \ ... \ ,n\}$) - объектами, а множество всех ее столбцов $\mathbf{\Omega} = \mathbf{C} = \{C_1, C_2, ..., C_n\}$ множеством примеров (или объектов) над универсумом **X**.

<u>Определение 9</u>

Матрицу $V = \mathbf{C}(V) \times \mathbf{R}(V)$, где $\mathbf{C}(V) \subseteq \mathbf{C}$, а $\mathbf{R}(V) \subseteq \mathbf{R}$, кроме того:

а) в $\mathbf{R}(V)$ найдется строка R_o, такая, что перенумеровав столбцы в множестве $\mathbf{C}(V)$ в соответствии с упорядочением элементов в строке R_o, мы получим, что в каждой из оставшихся строк множества $\mathbf{R}(V)$ значения x_{ij} изменяются монотонно вдоль множества столбцов $\mathbf{C}(V)$;

б) если x_{ol} представляет собой минимальное, а x_{or}- максимальное значение параметра R_o на множестве столбцов $\mathbf{C}(V)$, то неверно, что в матрице **X** найдется такой столбец $C_o \notin \mathbf{C}(V)$, что $x_{ol} \le x_{oo} \le x_{or}$;

в) $\mathbf{R}(V)$ есть максимальное по вложению множество строк, удовлетворяющее условиям а) и б) при фиксированном множестве $\mathbf{C}(V)$,

будем называть ***монотонной матрицей*** (***м-матрицей***) в **X**, а строку R_o - осью монотонности м-матрицы V.

<u>Определение 10</u>

Каждую максимальную по вложению (т.е. по числу) столбцов м-матрицу мы будем называть ***м-башней*** в универсуме **X**.

o

Пусть $V = \mathbf{C}_v \times \mathbf{R}_v = \{C_{i1}, C_{i2}, \ ... \ , C_{is}\} \times \{R_{i1}, R_{i2}, \ ... \ , R_{il}\}$ - м-матрица. Мы будем называть объекты $C_{i1}, C_{i2}, \ ... \ , C_{is}$ - ***родителями*** м-матрицы V, множество $\mathbf{C}_v = \{C_{i1}, C_{i2}, ..., C_{is}\}$ - ***орбитой*** м-матрицы V, а элементы множества $\mathbf{R}_v$ (т.е. имена строк $R_{i1}, R_{i2}, ..., R_{il}$) - ***образующими*** м-матрицы V. Множество всех м-матриц V, порождаемых над универсумом **X** из множества объектов $\mathbf{\Omega}$, мы будем обозначать посредством ***Мм***($\mathbf{\Omega}$). Его подмножество, содержащее лишь все м-башни, мы будем обозначать посредством ***Tow***($\mathbf{\Omega}$).

<u>Определение 11</u>

Шириной м-матрицы V будем называть мощность множества ее строк. ***Размером орбиты*** м-матрицы V будем называть мощность множества ее родителей (т.е. число образующих ее столбцов матрицы **X**).

o

Процедуры вычисления сходства объектов в рассматриваемом случае определяют оператор склейки м-матриц *CONCAT* и бинарная операция $\Pi\!\!\rightarrow$, предоставляющие возможности вычислять исчерпывающие сходства на множестве ***Мм***($\mathbf{\Omega}$) монотонных матриц над универсумом **X**.

<u>Определение 12</u>

На входе имеем две матрицы V_1 и V_2 в универсуме **X**. Результат работы алгоритма ***CONCAT*** - множество м-матриц $V_{1,2}$.

Алгоритм *CONCAT* задается следующим описанием:

Шаг 1. В множестве осей монотонности каждой из матриц V_1 и V_2 отметим общую ось R_j. (Если это невозможно, или все такие оси уже отмечены, переходим к Шагу 5).

Шаг 2. Построим множество $\mathbf{C}(V_1, V_2)$ всех столбцов из **X**, таких, что значения x_{ij} в строке R_j на столбцах из $\mathbf{C}(V_1, V_2)$ лежат между значениями x_{ij} в строке R_j на столбцах матриц V_1 и V_2.

Шаг 3. Построим множество $\mathbf{R}^j(V_1,V_2)$ - максимальное по вложению множество строк матрицы $\mathbf{X}$, таких, что

$$\mathbf{R}^j(V_1,V_2) \subseteq \mathbf{R}(V_1) \cap \mathbf{R}(V_2);$$

кроме того, значения x_{ij} в строках из $\mathbf{R}^j(V_1,V_2)$ изменяются ко-монотонно (см. определение м-матрицы) в смысле оси монотонности R_j вдоль столбцов из множества

$$\mathbf{C}^j(V_1,V_2) = \mathbf{C}(V_1) \cup \mathbf{C}(V_1,V_2) \cup \mathbf{C}(V_2).$$

Если в множестве $\mathbf{R}^j(V_1,V_2)$ окажется всего одна строка (это будет $\mathbf{R}^j(V_1,V_2)$) переходим к Шагу 1.

Шаг 4. Матрицу $\mathbf{V}^j{}_{1,2} = \mathbf{R}^j(V_1,V_2) \times \mathbf{C}^j(V_1,V_2)$ поместим в результирующее множество $V_{1,2}$ и перейдем к Шагу 1.

Шаг 5. Удалим из $V_{1,2}$ повторяющиеся элементы.

Шаг 6. СТОП.

o

Определение 13

Будем говорить, что примеры C_{i1} и C_{i2} локально сходны относительно матрицы $\mathbf{X}$, если среди порождаемых в $\mathbf{X}$ м-матриц найдется $V = \mathbf{R}(V) \times \mathbf{C}(V)$ такая, что $\mathbf{C}(V)=\{C_{i1},...,C_{i2}\}$.

o

Множество всех м-матриц в универсуме $\mathbf{X}$ обозначим через $\boldsymbol{\mu}(\mathbf{X})$. На множестве множеств м-матриц в $\mathbf{X}$ с помощью оператора *CONCAT* может быть определена бинарная операция ⧺, применяя которую к описаниям примеров (столбцам матриц $\mathbf{X}^+$ и $\mathbf{X}^-$)[7] как к одноэлементным множествам (*одностолбцовых*) м-матриц, можно найти:

а) максимальные по числу столбцов м-матрицы - *м-башни* в универсуме $\mathbf{X}$, а вместе с каждой из них:

б) явный вид множества параметров (имена строк соответствующих м-матриц), определяющих наличие (в случае анализа положительных примеров) или отсутствие (в случае анализа отрицательных примеров) изучаемого свойства;

в) интервалы допустимых изменений каждого из оказывающих влияние на наличие (или отсутствие) необходимого свойства параметров (диапазоны изменения числовых значений в каждой из строк соответствующей м-башни).

Таким путем, мы получаем возможность выделять как "*сопутствующим образом влияющие*" на поведение изучаемых объектов параметры, так и области их адекватных (в смысле наличия или отсутствия исследуемого свойства) изменений.

Итак, представленный алгоритм ***CONCAT*** перерабатывает подаваемую ему на вход пару $<V_1,V_2>$ элементов $\boldsymbol{\mu}(\mathbf{X})$ в множество V_{12}.

Лемма 12

Каждый элемент множества V_{12} - монотонная матрица из $\boldsymbol{\mu}(\mathbf{X})$.

Доказательство.

Для каждого элемента в множестве V_{12} его односвязность обеспечена действиями, выполняемыми на Шаге 2, а ко-монотонность значений в строках - действиями, выполняемыми на Шаге 3 в определении алгоритма ***CONCAT***.

o

Алгоритм CONCAT фактически определяет бинарный оператор $\cap_\mu$ склейки м-матриц, действующий при заданном "универсуме" $\mathbf{X}$ из $\boldsymbol{\mu}(\mathbf{X})\times\boldsymbol{\mu}(\mathbf{X})$ в множество всех подмножеств $2^{\boldsymbol{\mu}(\mathbf{X})}$. Ради наглядности изложения будем обозначать результат применения оператора $\cap_\mu$ к матрицам V_1 и V_2 посредством $V_1 \cap_\mu V_2$.

Лемма 13

Оператор $\cap_\mu$ удовлетворяет свойствам идемпотентности и коммутативности, т.е. для любых V_1 и V_2 имеют место:

(1) $V_1 \cap_\mu V_1 = V_1$

(2) $V_1 \cap_\mu V_2 = V_2 \cap_\mu V_1$

Доказательство.

Идемпотентность - очевидное свойство определения алгоритма ***CONCAT***. Коммутативность - очевидное следствие коммутативности теоретико-множественных операций $\cap$ и $\cup$, используемых при порождении всех $\mathbf{R}^j(V_1,V_2)$ и $\mathbf{C}^j(V_1,V_2)$ в определении алгоритма ***CONCAT***.

o

[7] Здесь мы полагаем, что $\mathbf{X}^+$ – множество позитивных, а $\mathbf{X}^-$ - множество негативных примеров для порождения (из них) эмпирических зависимостей, а каждый из их столбцов - примеров или контрпримеров - *по определению* считается м-матрицей.

Лемма 14

Пусть $V_1 \cap_\mu V_2 = V_{12}$, а $\mathbf{R}(V_1)$ и $\mathbf{R}(V_2)$ - множества строк соответствующих м-матриц V_1 и V_2. Тогда

$$|V_{12}| \le \min(|\mathbf{R}(V_1)|,|\mathbf{R}(V_2)|).$$

Доказательство.

Легко видеть, что число различных вариантов склейки определяется числом общих осей монотонности склеиваемых матриц.

о

Определим на множествах монотонных матриц из $\boldsymbol{\mu}(\mathbf{X})$ бинарную операцию $\Pi\!\!\!\rightarrow$ *локального относительного суперсходства*, действующую из $2^{\mu(X)}$ х $2^{\mu(X)}$ в $2^{\mu(X)}$.

Пусть $\mathbf{A} = \{A_1,..., A_2\}$ и $\mathbf{B} = \{B_1,...,B_2\}$ - произвольные множества монотонных матриц из $2^{\mu(X)}$:

$$\mathbf{A} \in 2^{\mu(X)} \text{ и } \mathbf{B} \in 2^{\mu(X)}.$$

Определение 14

Определим $\mathbf{D} = \mathbf{A} \;\Pi\!\!\!\rightarrow \mathbf{B}$ как "покомпонентное соединение" множеств матриц **A** и **B**:

$$\mathbf{D} = \mathbf{A} \;\Pi\!\!\!\rightarrow \mathbf{B} =$$
$$= \{A_1\cap_\mu B_1,..., A_1\cap_\mu B_2, \ldots , A_2\cap_\mu B_1,..., A_2\cap_\mu B_2\}$$

о

Считая, что множеству $2^{\mu(X)}$, по определению, принадлежит каждый столбец матрицы **X**, определим подмножество ***Dom***(**X**) множества $2^{\mu(X)}$ следующим образом:

1) каждый столбец матрицы **X** как одноэлементное множество монотонных матриц входит в ***Dom***(**X**);

2) если $\mathbf{A}\in$***Dom***(**X**) и $\mathbf{B}\in$***Dom***(**X**), то
$\mathbf{A} \;\Pi\!\!\!\rightarrow \mathbf{B} = \mathbf{D} \in$ ***Dom***(**X**).

3) других элементов в ***Dom***(**X**) нет.

Свойства операции $\Pi\!\!\!\rightarrow$ определяет

Теорема 8

Алгебра $\wp$ = <***Dom***(**X**),$\Pi\!\!\!\rightarrow$> есть нижняя полурешетка.

Доказательство.

1. Коммутативность.

Имея $\mathbf{A} = \{A_1,..., A_2\}$ и $\mathbf{B} = \{B_1,...,B_2\}$, вычислим

$$\mathbf{A} \;\Pi\!\!\!\rightarrow \mathbf{B} =$$
$$= \{A_1,..., A_2\} \;\Pi\!\!\!\rightarrow \{B_1,...,B_2\} =$$
$$= \{A_1\cap_\mu B_1,..., A_1\cap_\mu B_2, \ldots , A_2\cap_\mu B_1,..., A_2\cap_\mu B_2\} =$$ [8]
$$= \{B_1\cap_\mu A_1,..., B_1\cap_\mu A_2, \ldots , B_2\cap_\mu A_1,..., B_2\cap_\mu A_2\} =$$
$$= \mathbf{B} \;\Pi\!\!\!\rightarrow \mathbf{A}\ .$$

2. Ассоциативность.

Рассмотрим $\cap_\mu|_i$ - проекцию оператора $\cap_\mu$ на строку R_i исходной матрицы-"универсума" **X**, задаваемую следующим образом:

> $\cap_\mu|_i$ склеивает (порождая в виде результата не совпадающую с $\varnothing$ м-матрицу) только по оси монотонности R_i лишь такие м-матрицы V_1 и V_2, для которых R_i - общая ось монотонности.

Легко видеть, что $\cap_\mu|_i$ - ассоциативная операция: последовательность вычисления $(V_1 \cap_\mu|_i V_2) \cap_\mu|_i V_3$ дает (при склейке вдоль одной и той же оси R_i для всех трех матриц V_1,V_2 и V_3, благодаря ассоциативности теоретико-множественных операций $\cap$ и $\cup$), тот же результат, что и последовательность вычисления $V_1 \cap_\mu|_i (V_2 \cap_\mu|_i V_3)$. В свою очередь, ассоциативность $\Pi\!\!\!\rightarrow$ естественным образом следует из соотношения

$$(V_1 \cap_\mu V_2) = \bigcup_{R_i \in \mathbf{R}(\mathbf{X})} (V_1 \cap_\mu|_i V_2)$$ [9]

и ассоциативности теоретико-множественной операции $\cup$.

3. Идемпотентность.

Докажем необходимое индукцией по сложности $\mathbf{A}\in$***Dom***(**X**) (т.е. учтем, что каждый элемент из ***Dom***(**X**) порожден посредством операции $\Pi\!\!\!\rightarrow$ из "столбцов" матрицы **X**).

1) Если $\mathbf{A} = \{A\}$ - одноэлементное множество матриц и A - матрица, порождаемая одним из столбцов матрицы **X**, то, очевидно, что из идемпотентности оператора $\cap_\mu$ (**Лемма 13**) имеем:

$$\{A\} \;\Pi\!\!\!\rightarrow \{A\} = \{A\}\ .$$

2) По определению ***Dom***(**X**), для каждого $\mathbf{B}\in$***Dom***(**X**) найдутся $\mathbf{B_1}$ и $\mathbf{B_2}$ из ***Dom***(**X**), такие, что $\mathbf{B} = \mathbf{B_1} \;\Pi\!\!\!\rightarrow \mathbf{B_2}$. Будем считать, что

$$\mathbf{B_1} \;\Pi\!\!\!\rightarrow \mathbf{B_1} = \mathbf{B_1} \text{ и } \mathbf{B_2} \;\Pi\!\!\!\rightarrow \mathbf{B_2} = \mathbf{B_2}\ ,$$

тогда имеем:

[8] В силу коммутативности оператора $\cap_\mu$.
[9] Т.е. объединением по всем возможным осям монотонности.

$$\mathbf{B} \sqcap \mathbf{B} = (\mathbf{B}_1 \sqcap \mathbf{B}_2) \sqcap (\mathbf{B}_1 \sqcap \mathbf{B}_2) =$$
$$=^{10} \mathbf{B}_1 \sqcap \mathbf{B}_2 \sqcap \mathbf{B}_1 \sqcap \mathbf{B}_2 =$$
$$=^{11} \mathbf{B}_1 \sqcap \mathbf{B}_1 \sqcap \mathbf{B}_2 \sqcap \mathbf{B}_2 =$$
$$=^{12} (\mathbf{B}_1 \sqcap \mathbf{B}_1) \sqcap (\mathbf{B}_2 \sqcap \mathbf{B}_2) =$$
$$= \mathbf{B}_1 \sqcap \mathbf{B}_2 = \mathbf{B}\ .$$

o

В случае операций с не принадлежащими ***Dom*(X)** элементами нарушение рассматриваемого свойства (т.е. невозможность распространить свойство идемпотентности $\sqcap$ с ***Dom*(X)** на все множество $2^{\mu(X)}$) иллюстрирует

Пример 1.

Пусть $\mathbf{A} = \{\mathbf{A_1}, \mathbf{A_2}\}$, где $\mathbf{A_1}$ есть матрица

	C_1	C_2	C_3
R_1	1	2	3
R_2	11	12	13

и $\mathbf{A_2}$ есть матрица

	C_4	C_5	C_6
R_1	4	5	6
R_2	14	15	16

В таком случае, учитывая $\mathbf{A_3}$:

	C_1	C_2	C_3	C_4	C_5	C_6
R_1	1	2	3	4	5	6
R_2	11	12	13	14	15	16

легко видеть, что

$$\mathbf{A} \sqcap \mathbf{A} = \{\mathbf{A_1}, \mathbf{A_2}, \mathbf{A_3}\} \neq \mathbf{A}.$$

o

7. Работа с числовыми векторами: оценки сложности вычислений

Теперь мы имеем возможность обратиться собственно к анализу комбинаторных свойств элементов множеств ***Мм*(Ω)** и ***Tow*(Ω)**.

Определение 15

Задачей ***ПРОИЗВОЛЬНАЯ М-БАШНЯ*** (***М-БАШНЯ***) будем называть следующую задачу:

Дано: числовая матрица **X**.

Найти: какую-либо м-башню[13] в множестве ***Мм*(Ω)**

o

Точное представление о комбинаторной природе множества ***Tow*(Ω)** дают **Теоремы 9** и **10.**

Теорема 9

Задача ***ПРОИЗВОЛЬНАЯ М-БАШНЯ*** при обработке числовых матриц лежит в классе полиномиально разрешимых задач.

Доказательство.

Легко видеть, что любой элемент множества ***Tow*(Ω)** может быть получен посредством следующей процедуры:

- выберем в универсуме **X** минор, порожденный пересечением ровно двух строк и двух столбцов (т.е. образованный четверкой чисел);
- будем перебирать последовательно все оставшиеся столбцы (добавляя их к выбранному минору), стремясь пополнить имеющееся множество столбцов, сохранив при этом монотонность изменения числовых значений вдоль строк расширяемого таким образом минора. Получив первый из столбцов, сохраняющий ко-монотонность изменения числовых значений вдоль строк минора, продолжим расширение до состояния нерасширяемости (по столбцам), затем
- будем перебирать последовательно все оставшиеся строки матрицы **X**, повторив и здесь (по аналогии со столбцами) процедуру пополнения выбранного нами минора.

Результат применения процедуры для каждого исходного минора - некоторая м-башня (содержащая исходный минор), либо доказательство того факта, что исходный минор не может быть расширен в универсуме **X** до какой-либо м-башни.

Учитывая, что число “стартовых” миноров для представленной процедуры в универсуме **X** не превосходит величины

$$C^n_2 \times C^m_2 \times (n-2) \times (m-2) \leq$$
$$= O\left(\frac{n(n-1)(n-2)m(m-1)(m-2)}{4}\right) = O(n^3 \times m^3),$$

а пополнение каждого из “стартовых” миноров (до завершения представленной процедуры) может быть осуществлено однократным про-

---

[10] По ассоциативности $\sqcap$.

[11] По коммутативности $\sqcap$.

[12] По исходному предположению.

[13] При этом мы будем рассматривать лишь нетривиальные случаи, т.е. м-башни, имеющие не менее трех родителей.

смотром множеств строк и столбцов матрицы **X**, мы приходим к искомому утверждению.

оо

<u>Теорема 10</u>

Задача ***ПРОИЗВОЛЬНАЯ М-БАШНЯ*** при обработке числовых матриц перечислительно полна:

М-БАШНЯ $\in$ ***#PC***

<u>Доказательство.</u>

Идея доказательства - продемонстрировать полиномиальную сводимость известной (например, [9]) переборной задачи ***МОНОТОННАЯ ВЫПОЛНИМОСТЬ*** к рассматриваемой нами задаче ***ПРОИЗВОЛЬНАЯ М-БАШНЯ***. Мы воспользуемся той же схемой сведения, которую мы использовали при доказательстве аналогичных утверждений при работе с данными в виде множеств: мы будем перечислять нули соответствующей монотонной булевской функции **Φ**, представленной в виде 2-КНФ, и кодировать их с помощью м-башен в некотором специально построенном универсуме (числовой матрице).

Теперь более подробно. Пусть **Φ** имеет вид

$$\mathbf{\Phi}(x_1,x_2, \dots ,x_n) = \mathbf{\Phi_1} \ \& \ \mathbf{\Phi_2} \ \& \dots \& \ \mathbf{\Phi_s} =$$

$$= (x_{11} \vee x_{12}) \ \& \ (x_{21} \vee x_{22}) \ \& \dots \& \ (x_{s1} \vee x_{s2}).$$

Не теряя общности, будем считать, что в представлении функции **Φ** все конъюнкты $\mathbf{\Phi_i}$ различны (в противном случае функция **Φ** может быть приведена к нужному нам виду полиномиально сложной процедурой - сортировкой образующих ее конъюнктов). Положим, что универсум для задачи о числе исчерпывающих пересечений множеств $\mathbf{U_\Phi} = \{a_1,a_2, \dots ,a_n\}$, а множество объектов над выбранным универсумом $\mathbf{\Omega_\Phi} = \{A_1,A_2, \dots ,A_m,A_{m+1}\}$ определяется условиями:

$$A_\mathbf{j} = \{a_1,a_2, \dots ,a_n\} \setminus \{a_{i1},a_{i2}\} \quad \text{для } \mathbf{i} \in \{\mathbf{1,2, \dots ,n}\} \text{ и } \mathbf{j} \in \{\mathbf{1,2, \dots ,s}\},$$

$$A_{\mathbf{s+1}} = \{a_1,a_2, \dots ,a_n\}.$$

Фактически, каждый из объектов A_i (для $1 \le \mathbf{i} \le \mathbf{s}$) сопоставляется нулю соответствующего конъюнкта $\mathbf{\Phi_i}$ из 2-КНФ представления функции **Φ**.

Множество исчерпывающих сходств объектов из $\mathbf{\Omega_\Phi}$ взаимно однозначным образом перечисляет все нули функции **Φ**. Теперь рассмотрим матрицу $\mathbf{X_\Phi}$ образованную n строками (по числу переменных функции **Φ**) и s+1 столбцами (по числу объектов в множестве $\mathbf{\Omega_\Phi}$). Положим в $\mathbf{X_\Phi}$ элемент, стоящий на пересечении строки **i** и столбца **j**, равным 1 тогда и только тогда, когда образующая a_i из универсума $\mathbf{U_\Phi}$ входит в объект $A_\mathbf{j}$ из множества $\mathbf{\Omega_\Phi}$. Все остальные элементы матрицы $\mathbf{X_\Phi}$ - нули.

Легко видеть, что каждому исчерпывающему сходству объектов из $\mathbf{\Omega_\Phi}$ соответствует некоторая м-башня в матрице $\mathbf{X_\Phi}$. Действительно, имея некоторое исчерпывающее сходство $V_\mathbf{\Phi}$, сформированное образующими $a_{\mathbf{i1}}, \dots , a_{\mathbf{i2}}$ из родителей $A_{\mathbf{j1}}, \dots , A_{\mathbf{j2}}$, мы можем перенумеровать строки и столбцы матрицы $\mathbf{X_\Phi}$ так, что представляющий $V_\mathbf{\Phi}$ минор будет находиться в левом верхнем углу матрицы $\mathbf{X_\Phi}$. Зафиксируем одну из образующих $V_\mathbf{\Phi}$ строк как ось монотонности порождаемой нами м-башни. Теперь просмотром оставшихся столбцов матрицы $\mathbf{X_\Phi}$ мы можем построить единственное (ось монотонности фиксирована!) расширение множества столбцов минора $V_\mathbf{\Phi}$ (за счет учета в м-башне новых элементов, находящихся в соответствующих строках), где числовые значения в строках расширенного новыми столбцами минора $V'_\mathbf{\Phi}$ не возрастают. Таким образом, каждому исчерпывающему сходству из $\boldsymbol{GC}(\mathbf{U_\Phi},\mathbf{\Omega_\Phi})$ при фиксированной оси монотонности) сопоставляется единственным образом некоторая м-башня из множества $\boldsymbol{Tow}(\mathbf{X_\Phi})$.

Теперь нам осталось убедиться, что мы в состоянии (полиномиально быстро!) пересчитать количество м-башен сопоставляемых каждому исчерпывающему сходству (пересечению) объектов из $\mathbf{\Omega_\Phi}$. Обратим внимание на то, что число м-башен, сопоставляемых каждому исчерпывающему сходству (пересечению) из $\boldsymbol{GC}(\mathbf{U_\Phi},\mathbf{\Omega_\Phi})$ определяется числом осей монотонности в миноре, представляющем собственно анализируемое исчерпывающее сходству, а таких осей ровно столько, сколько строк в этом миноре (совпадающих осей нет, так как мы договорились, что среди конъюнктов в 2-КНФ представлении функции **Φ** нет повторяющихся и, следовательно, среди “примитивных” нулей - и строк матрицы $\mathbf{X_\Phi}$ - нет повторяющихся). Остается заметить, что каждому минору, содержащему q строк, сопоставляется ровно q экземпляров м-башен, а число таких миноров есть число комбинаций по q штук из “примитивных” нулей функции **Φ**, т.е.

$$N_q = C^s_q$$

Учитывая, что всего осей монотонности ровно столько, сколько строк в матрице $\mathbf{X}_{\Phi}$, и принимая во внимание, что матрица $\mathbf{X}_{\Phi}$ порождается по функции **Ф** полиномиально быстро (столь же быстро, сколь и универсум $\mathbf{U}_{\Phi}$ и множество объектов $\mathbf{\Omega}_{\Phi}$), мы получаем искомое утверждение.

оо

Экспоненциально быстрый в (общем случае) рост числа элементов множества ***Tow***(**X**) м-башен в универсуме **X** обусловлен возможными повторениями чисел в строках матрицы **X**. Исключив подобные повторения, мы получим быстро обозримый частный случай, представление о котором дает

Теорема 11

Задача ***ПРОИЗВОЛЬНАЯ М-БАШНЯ*** при обработке числовых матриц в случае отсутствия в строках матрицы-универсума **X** одинаковых элементов имеет полиномиально большое множество решений.

Доказательство.

Легко видеть, что в рассматриваемой ситуации каждая из осей монотонности имеет ровно один вариант упорядочения как по возрастанию, так и по убыванию. Представленная в доказательстве **Теорема 8** процедура “выращивания” м-башен из 4-х-элементных миноров позволяет породить любой элемент множества ***Tow***(**X**). Однозначный вариант упорядочения элементов каждой из строк матрицы **X** делает справедливой оценку сверху

$$C^n_2 \times C^m_2 \times (n-2) \times (m-2) =$$

$$=\text{Const} \times \left(\frac{n(n-1)(n-2)m(m-1)(m-2)}{4}\right)=O(n^3 \times m^3) \geq N_X$$

для числа N_X искомых м-башен.

оо

Теперь, по аналогии с описанием комбинаторных свойств соответствующих переборных задач в случае обработки ранее рассмотренных типов данных, мы имеем возможность сформулировать

Определение 16

1. Задача ***ВСЕ М-БАШНИ МАКСИМАЛЬНОЙ ШИРИНЫ*** при обработке числовых матриц определяется условиями:

   Дано: универсум **X** (а вместе с ним и множество объектов **Ω**).

   Найти: множество всех максимальных по вложению образующих их множеств строк элементов в ***Tow***(**X**).

2. Задача ***ВСЕ М-БАШНИ МИНИМАЛЬНОЙ ШИРИНЫ*** при обработке числовых матриц определяется условиями:

   Дано: универсум **X** (а вместе с ним и множество объектов **Ω**).

   Найти: множество всех минимальных по вложению образующих их множеств строк элементов в ***Tow***(**X**).

3. Задача ***ВСЕ М-БАШНИ С ОРБИТОЙ МАКСИМАЛЬНОГО РАЗМЕРА*** при обработке числовых матриц определяется условиями:

   Дано: универсум **X** (а вместе с ним и множество объектов **Ω**).

   Найти: множество всех элементов в ***Tow***(**X**), характеризуемых максимальным по вложению размером орбит.

4. Задача ***ВСЕ М-БАШНИ С ОРБИТОЙ МИНИМАЛЬНОГО РАЗМЕРА*** при обработке числовых матриц определяется условиями:

   Дано: универсум **X** (а вместе с ним и множество объектов **Ω**).

   Найти: множество всех элементов в ***Tow***(**X**), характеризуемых минимальным по вложению размером орбит.

5. Задача ***М-БАШНЯ ШИРИНЫ НЕ МЕНЕЕ k*** при обработке числовых матриц определяется условиями:

   Дано: натуральное число ***k***, универсум **X** (а вместе с ним и множество объектов **Ω**).

   Найти: элемент множества ***Tow***(**X**), образованный не менее ***k*** строками из множества $\mathbf{R}_X$.

6. Задача ***М-БАШНЯ ШИРИНЫ НЕ БОЛЕЕ k*** при обработке числовых матриц определяется условиями:

   Дано: натуральное число ***k***, универсум **X** (а вместе с ним и множество объектов **Ω**).

   Найти: элемент множества ***Tow***(**X**), образованный не более ***k*** строками из множества $\mathbf{R}_X$.

7. Задача ***М-БАШНЯ С ОРБИТОЙ РАЗМЕРА НЕ МЕНЕЕ k*** при обработке числовых матриц определяется условиями:

   Дано: натуральное число ***k***, универсум **X** (а вместе с ним и множество объектов **Ω**).

<u>Найти</u>: элемент множества ***Tow***(**X**), орбита которого содержит не менее ***k*** элементов из множества **Ω**.

8. Задача ***М-БАШНЯ С ОРБИТОЙ РАЗМЕРА НЕ БОЛЕЕ k*** при обработке числовых матриц определяется условиями:

<u>Дано</u>: натуральное число ***k***, универсум **X** (а вместе с ним и множество объектов **Ω**).

<u>Найти</u>: элемент множества ***Tow***(**X**), орбита которого содержит не более ***k*** элементов из множества **Ω**.

9. Задача ***М-БАШНЯ ШИРИНЫ РОВНО k*** при обработке числовых матриц определяется условиями:

<u>Дано</u>: натуральное число ***k***, универсум **X** (а вместе с ним и множество объектов **Ω**).

<u>Найти</u>: элемент множества ***Tow***(**X**), образованный ровно ***k*** строками из множества $\mathbf{R}_\mathbf{X}$.

10. Задача ***М-БАШНЯ С ОРБИТОЙ РАЗМЕРА РОВНО k*** при обработке числовых матриц определяется условиями:

<u>Дано</u>: натуральное число ***k***, универсум **X** (а вместе с ним и множество объектов **Ω**).

<u>Найти</u>: элемент множества ***Tow***(**X**), орбита которого содержит ровно ***k*** элементов из множества **Ω**.

о

Вычислительную сложность порождения границ каждого множества м-башен характеризуют **Теоремы 12-13** и **Следствия 16-17**:

<u>Теорема 12</u>

Задача ***ВСЕ М-БАШНИ МАКСИМАЛЬНОЙ ШИРИНЫ*** при обработке числовых матриц полиномиально разрешима (т.е. существует алгоритм, порождающий ее решение в ходе вычислений, объем которых ограничен сверху полиномом от характерных размеров входных данных).

<u>Доказательство.</u>

Легко видеть, что каждая м-башня максимальной ширины всегда содержит тройку находящихся внутри нее рядом (в смысле упорядочения вдоль некоторой оси монотонности) столбцов. Поиск такой тройки столбцов и позволит нам выделить (достраивая добавлением новых столбцов найденные тройки до максимального возможного количества объектов) все искомые м-башни.

Рассмотрим множество троек столбцов из **Ω**. Оставим в нем лишь те тройки, которые порождают (на подмножествах своих строк) некоторые м-матрицы. Нетрудно убедиться, что таких троек не более

$$C^m{}_3 = \text{m(m-1)(m-2)\textbackslash 6}$$

штук. Легко видеть, что для каждой такой тройки столбцов можно предложить не более (3!) x m = 6m упорядочений (вариантов выбора оси монотонности) элементов в строках. К каждому из не более чем 6m вариантов упорядочения каждой из отобранных троек столбцов добавим (перебором по оставшимся m-3 столбцам) по одному столбцу. Построим на каждой из построенных четверок столбцов все возможные упорядочения (их не более 4!xm = 24m), и оставим среди полученных четверок лишь те, которые представляют собой расширение (приписыванием четвертого столбца справа или слева) одной из ранее построенных троек столбцов, сохраняющее характер упорядочений (там, где оно имело место) на строках выбранной тройки столбцов. (Другими словами, выделим на тройках столбцов те, которые порождают некоторые м-матрицы, убедившись в том, что это именно м-матрицы, проанализировав все возможные расширения троек столбцов до четверок столбцов). Нам осталось среди отобранных троек столбцов убрать те, для которых множество строк содержащейся в тройке столбцов м-матрицы вкладывается в множество строк содержащейся в некоторой другой тройке столбцов м-матрицы (т.е. удалить “немаксимальные” элементы). Процедура проверки вложимости имеет характер сортировки и, как нетрудно убедиться, также полиномиально реализуема.

оо

<u>Следствие 16</u>

Задача ***ВСЕ М-БАШНИ С ОРБИТОЙ МИНИМАЛЬНОГО РАЗМЕРА*** при обработке числовых матриц полиномиально разрешима.

<u>Доказательство.</u>

Легко видеть, что каждая м-башня с орбитой минимального размера имеет максимальное по вложению множество строк.

о

<u>Теорема 13</u>

Задача ***ВСЕ М-БАШНИ МИНИМАЛЬНОЙ ШИРИНЫ*** при обработке числовых матриц полиномиально разрешима.

<u>Доказательство.</u>

Для доказательства достаточно рассмотреть в множестве троек столбцов, использованных в доказательстве **Теорема 12**, минимальные по вложимости множества строк м-матрицы.

оо

Следствие 17

Задача ***ВСЕ М-БАШНИ С ОРБИТОЙ МАКСИМАЛЬНОГО РАЗМЕРА*** при обработке числовых матриц полиномиально разрешима.

Доказательство.

Легко видеть, что каждая м-башня с орбитой максимального размера имеет минимальное по вложению множество строк.

о

Вычислительную сложность порождения фрагментов каждого множества м-башен, характеризуемых числовыми ограничениями на их размеры, представляют теоремы 14 и 15.

Теоремы 14

Следующие переборные задачи при обработке числовых матриц полиномиально разрешимы:

1. ***М-БАШНЯ ШИРИНЫ НЕ МЕНЕЕ k***
2. ***М-БАШНЯ ШИРИНЫ НЕ БОЛЕЕ k***
3. ***М-БАШНЯ С ОРБИТОЙ РАЗМЕРА НЕ МЕНЕЕ k***
4. ***М-БАШНЯ С ОРБИТОЙ РАЗМЕРА НЕ БОЛЕЕ k***
5. ***М-БАШНЯ ШИРИНЫ РОВНО k***

Доказательство.

Декларируемые свойства задач 1, 2, 5 определяются, как и при доказательстве **Теоремы 12**, наличием в м-башне, обладающей искомыми свойствами, трех расположенных рядом столбцов, которые содержат м-матрицу (трехстолбцовую!) с искомыми свойствами. Таким образом, полиномиальный алгоритм перебора трех- и четырехстолбцовых фрагментов матрицы-универсума, использованный при доказательстве **Теоремы 12**, способен привести нас к цели и в рассматриваемом случае.

Декларируемые свойства задач 3 и 4 определяются анализом множества м-башен с орбитами максимального и минимального размера. Отсутствие среди м-башен с орбитами максимального размера такой, которая была бы образована не менее ***k*** объектами (а это по **Следствию 17** может быть выяснено полиномиально быстро) гарантирует отсутствие таковых вообще. Аналогично, отсутствие среди м-башен с орбитами минимального размера такой, которая была бы образована не более ***k*** объектами (а это по **Следствию 16** может быть выяснено полиномиально быстро) также гарантирует отсутствие таковых вообще.

оо

Теорема 15

Задача ***М-БАШНЯ С ОРБИТОЙ РАЗМЕРА РОВНО k*** при обработке числовых матриц принадлежит классу **NP**-полных задач.

Доказательство.

Прежде всего убедимся, что по заданным произвольным образом матрице-универсуме X и некоторой числовой матрице V можно полиномиально быстро проверить, является ли V м-башней с орбитой размера ровно ***k*** в универсуме X. Действительно, для этого достаточно проверить

- вложимость столбцов матрицы V (как подстолбцов) в столбцы матрицы X;
- возможность пополнить (последовательным перебором оставшихся строк матрицы X) матрицу V до м-матрицы, содержащей большее чем в V число строк;
- возможность пополнить (последовательным перебором оставшихся столбцов матрицы X) матрицу V до м-матрицы, содержащей большее число столбцов.

Далее в доказательстве мы воспользуемся тем же механизмом сводимости, который был нами задействован при доказательстве **Теоремы 5**. Так же, как и ранее мы воспользуемся сводимостью задачи ***ВЫПОЛНИМОСТЬ БУЛЕВСКОЙ ФУНКЦИИ*** к задаче ***ИСЧЕРПЫВАЮЩЕЕ СХОДСТВО РАЗМЕРА РОВНО k***. Далее, как и при доказательстве **Теоремы 10**, матрицу-универсум $\mathbf{X_\Phi}$ мы построим из строк, сопоставленных всем отдельным переменным функции **Φ** и их отрицаниям, и столбцов, сопоставленных образующим универсума $\mathbf{U_\Phi}$ (см. образующие множества $\omega = \{Y_1, Y_2, \ldots, Y_{2n}\}$ из доказательства **Теоремы 5**), порождаемого по виду исходной булевской функции **Φ**, заданной нам в форме 3-КНФ. Все перечисленные строки и столбцы в матрице $\mathbf{X_\Phi}$ заполнены 0 (в случае, когда соответствующая образующая не входит в соответствующее множество образующих, представленное соответствующей строкой матрицы $\mathbf{X_\Phi}$) и 1 (в случае, когда соответствующая образующая не входит в соответствующее множество образующих, представленное соответствующей строкой матрицы $\mathbf{X_\Phi}$). Дополнительная, 2n+1-я строка универсума $\mathbf{X}_{\Phi}$ образована по следующему закону:

- во всех столбцах, соответствующих образующим, которые не войдут в возможные исчерпывающие пересечения размера ровно n, находятся 0;
- в оставшихся 2n столбцах по строке номер 2n+1 расставлены числа 1,2, ... ,2n-2,2n-1, 2n-1 (в первом и втором - 1 и 2; в третьем и четвертом - 3 и 4, и т.д. до соответствующих значений в предпоследней паре столбцов и чисел 2n-1 в двух последних столбцах).

Легко видеть, что перенумеровав (что может быть сопоставлено выбору подходящей оси монотонности и последующему построению необходимой м-башни) соответствующим образом строки и столбцы матрицы $\mathbf{X}_{\Phi}$, можно каждое исчерпывающее пересечение размера ровно n представить минором размера n(столбцов) на n+1(строк), где в строке номер n+1 монотонное возрастание значений от 1 до 2n-1, прерывается 0 на первом же столбце с номером более n+1. (В столбце номер n+1 в строках с номерами меньшими чем 2n+1 мы найдем нули и единицы, что не изменит направления монотонности “невозрастания” в каждой из этих строк. В строке же номер 2n+1 в этом столбце будет стоять число 2n-1, превосходящее любое из чисел из множества {0,1,2, ... , 2n-2}, расположенных в этой строке на пересечении со столбцами с номером более n+1). Таким образом, каждому исчерпывающему пересечению размера ровно n в универсуме $\mathbf{U}_{\Phi}$ взаимно однозначным образом сопоставляется м-башня, построенная на n строках и n+1 столбце матрицы-универсума $\mathbf{X}_{\Phi}$.

Для завершения доказательства **Теоремы 15** нам достаточно убедиться, что матрица-универсум $\mathbf{X}_{\Phi}$ строится полиномиально быстро по заданной функции **Φ** (что действительно справедливо, т.к. матрица-универсум $\mathbf{X}_{\Phi}$ всего одной строкой отличается от множества объектов над универсумом $\mathbf{U}_{\Phi}$, для которых исчерпывающие пересечения необходимого размера перечисляют выполняющие функцию **Φ** наборы значений булевских переменных).

oo

Заключение

Завершая обзор оценок сложности вычислений, характеризующих обсуждаемую нами схему ИАД, обратим внимание на некоторые отличия ситуации с обработкой данных в виде векторов числовых значений переменных (в частности – результат **Теоремы 14** относительно *полиномиальной* разрешимости задачи ***М-БАШНЯ ШИРИНЫ РОВНО k***). Здесь существенную роль играет уже отмечавшийся нами выше эффект учета *контекста*, в рамках которого и производится вычисление соответствующих сходств: сходство объектов (столбцов исходной матрицы-универсума **X**) определяется наличием «*пути*» (последовательности промежуточных объектов, на соответствующих значениях параметров которых сохраняется условие *ко-монотонности* [14]изменений их числовых значений). Именно это дополнительное обстоятельство – учет *контекста* – дает возможность избежать здесь экспоненциально быстрого комбинаторного роста требующих анализа вариантов. Однако, и здесь (см. обозначенный **Теоремой 15** результат о **NP**–полноте задачи ***М-БАШНЯ С ОРБИТОЙ РАЗМЕРА РОВНО k***) уже знакомая нам по работе с другими, ранее рассмотренными, типами данных ситуация трудно-разрешимости задачи о порождении класса эквивалентности заданного размера вновь оказывается актуальной характеристикой требующих решения комбинаторных проблем.

Таким образом, в ситуации возникновения подробно рассмотренных нами трудностей с вычислительной сложностью соответствующих алгоритмов ИАД оказывается весьма полезно для применения в реальных индустриальных приложениях иметь соответствующие «процедурные» возможности

- быстро выделять эффективно разрешимые подклассы решаемых переборных задач восстановления зависимостей, а также
- реализовать (в том или иной виде формализованные) *приближенные* вычисления.

Именно такой путь предложен в работах [10-12], где обсуждаются возможности специальной проблемно-ориентированной техники оптимизации перебора в рамках так называемого *приближенного* ДСМ-метода, позволяющего целенаправленно управлять объемами вычис-

[14] Напомним, что в рамках выбранной нами семантики именно этим условием характеризуется Миллевское (подробнее [5]) условие сопутствующих изменений наблюдаемых «каузальных» параметров.

лений, которые необходимы для реализации ДСМ-рассуждений. В основе предлагаемого механизма оптимизации ИАД-расчетов лежит специальным образом организованная технология гибкой настройки и управления «навигацией» в множестве соответствующих ДСМ-сходств (целенаправленное управление перебором элементов множества потенциально порождаемых в каждом конкретном случае ДСМ-гипотез). Именно таким образом и сформирована технология восстановления эмпирических зависимостей из данных, позволяющая вести средствами ДСМ-метода ИАД в том числе и на больших массивах.

Литература

1. Забежайло М.И., Синякова Е.В. К вопросу об «интеллектуальности» интеллектуального анализа данных. – Научно-техническая информация. Сер. 2, Информационные процессы и системы. - 2014. - № 3. - С. 1-9.
2. Turing A. Computing Machinery and Intelligence. - Mind, vol. LIX, № 236, October 1950, pp. 433—460.
3. Эшби У.Р. Введение в кибернетику. – М.: Иностранная литература, 1959. – 432 С. (http://pcp.vub.ac.be/ASHBBOOK.html)
4. Автоматическое порождение гипотез в интеллектуальных системах. Под общ. ред. В.К. Финна. – М.: Книжный дом «Либроком», 2009. – 526 С.
5. Милль Д.С. Система логики силлогистической и индуктивной. – М.; Книжное дело. - 1900. - 781 С.
6. Забежайло М.И. К проблеме расширения ДСМ-метода на данные с числовыми параметрами. – Научно-техническая информация. Сер.2. – 1992. – №11, С.11-21.
7. Забежайло М.И. К проблеме расширения ДСМ-метода на данные с числовыми параметрами II. – Научно-техническая информация. Сер.2. – 1993. – №2, С. 6-16.
8. Забежайло М.И. К проблеме формализации метода геологических аналогий. Известия РАН. Теория и системы управления. -1997. №2, С.151-164.
9. Valiant L.G. The complexity of enumeration and reliability problems. - SIAM J.Comput. - 8 (1979), N1. - Pp. 410-421.
10. Забежайло М.И. О некоторых возможностях управления перебором в ДСМ-методе // Искусственный интеллект и принятие решений. – 2014. – Часть I: № 1, С.95 -110.
11. Забежайло М.И. О некоторых возможностях управления перебором в ДСМ-методе // Искусственный интеллект и принятие решений. – 2014. – Часть II: № 3, С.3 - 21.
12. Забежайло М.И. Приближенный ДСМ-метода на примерах. – Научно-техническая информация. Сер.2. – 2014. – №10, С. 1-12.

Забежайло Михаил Иванович. Управляющий директор Центра прикладных исследований компьютерных сетей. Окончил Московский физико-технический институт в 1979 году. Кандидат физико-математических наук, старший научный сотрудник (доцент). Автор более 70 печатных работ. Область научных интересов: интеллектуальный анализ данных.
E-mail: mzabezhailo@arccn.ru