

О некоторых оценках сложности вычислений в ДСМ-методе. Часть I

Аннотация. Обсуждаются возможности организации интеллектуального анализа данных (ИАД), обеспечивающего экспертизу различных типов представления исходных данных (множеств признаков, графов, числовых векторов) однородными средствами их обработки. Рассматриваются некоторые процедурные особенности ДСМ-метода автоматического восстановления зависимостей из данных, позволяющие использовать его как платформу для организации ИАД. 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.

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

Введение

Интеллектуальный анализ данных (ИАД) [1, 2] сегодня позиционируется как одно из актуальных и быстро развивающихся направлений *computer science*. Приложения ИАД в области анализа сверхбольших коллекций эмпирических данных (так называемая проблематика *big data* – например [3]) привлекают особое внимание и исследователей (ориентированных на целый ряд интересных задач, возникающих в этой области науки и технологий), и, конечно же, так называемых «конечных пользователей» (т.е. тех, кому возможности получать в рассматриваемой области приемлемые решения соответствующих прикладных задач позволяют развивать бизнес – создавать новые инженерные решения, эффективно управлять крупными техническими или же финансовыми системами и т.п.).

Достаточно емкий (по встречаемости в практически важных приложениях) класс задач ИАД, ориентированных на восстановление эмпирических зависимостей, которые неявным образом представлены в тех или иных коллекциях экспериментальных данных, в самом общем виде можно описать следующей схемой (*ПС - прогноза свойств*):

Даны:

1) множество (примеров) *объектов*, обладающих целевыми *свойствами*, и множество (в исходном состоянии, – возможно, пустое) объектов, не обладающих целевыми свойствами – контрпримеров,

а также

2) *новый объект* (или же некоторое явным образом представленное множество таких объектов).

Требуется:

оценить *наличие* (или отсутствие) *целевых свойств* у нового объекта (новых объектов из заданного множества), т.е.

- дать соответствующий *прогноз* (о наличии целевых свойств) и

- предъявить *основания* (неоспоримые *аргументы*), позволяющие *принять* этот прогноз.

Формализованные постановки задач этого вида обычно используют вполне определенный набор математических моделей и алгоритмов. Если говорить о наиболее распространенных инструментах подобного типа, то в первую очередь, по-видимому, следует упомянуть модели и методы статистического анализа данных [4-6], машинного обучения [7, 8], распознава-

ния образов и автоматической классификации [9-11], нейронных сетей [12-13], а также «технику» генетических алгоритмов [14-15].

Однако опыт реализации в рамках ИАД перечисленных выше подходов демонстрирует существование, по крайней мере, трех существенных проблем, ограничивающих применение такого «инструментария» в практических (например, индустриальных) приложениях. В самом общем виде эти проблемы можно сформулировать следующим образом:

(а) Применением конкретной математической техники ИАД рассчитан некоторый результат. Какими аргументами в пользу *достаточности оснований для принятия результатов* выполненных вычислений мы располагаем? (Какими средствами контроля *достаточности оснований «доверять»* полученным результатам можно обосновать их «надежность»?)

(b) Как тот или иной математический инструментарий ИАД обеспечивает корректность обработки «малых» (например, статистически не значимых) выборок исходных данных?

(с) Как тот или иной математический инструментарий ИАД справляется с проблемами *вычислительной сложности* предлагаемых ему для решения переборных задач¹?

Принимая во внимание эти ограничения, придется разделить первые три (см. выше) и последующие два примера математических «техник» ИАД. Действительно, при использовании нейронных сетей или же генетических алгоритмов мы в общем случае не располагаем какими-либо внешними (т.е. не воспроизводящими собственно алгоритмику реализуемых вычислений) аргументами, позволяющими судить о причинах возникновения формируемых эмпирических зависимостей. (Фактически, здесь мы имеем дело лишь со способом вычислений. Ответы на вопросы вида

- **ПОЧЕМУ** имеет место конкретная эмпирическая зависимость?

- что является *причиной* ее возникновения?

оказываются за рамками «выразительных возможностей» используемой техники ИАД).

¹ Напомним, что в целом ряде практически важных случаев при реализации ИАД приходится сталкиваться не только с необходимостью оперировать достаточно большими объемами исходных данных (*big data* - !), но и с проблемой доказуемой трудно-разрешимости (например, NP-полноты и т.п.) соответствующих комбинаторных задач.

В свою очередь первые три из названных выше классов математических инструментов ИАД, предлагают свои - вполне «респектабельно выглядящие» - варианты ответа на эти вопросы (предлагая свои версии реконструкции из имеющихся эмпирических данных если уж не *причинных*, то, по крайней мере, *ассоциативных* зависимостей, например, - статистически значимых корреляций и т.п.). Тем не менее, здесь очевидным образом проявляется ряд ограничений (в части возможностей, доступных этим инструментам ИАД) в области ключевых аспектов проблематики (b) и (с). Так, например, вопрос о корректной применимости статистических методов анализа данных [4-6] при обработке малых выборок сложно структурированных объектов остается открытым и сегодня. В свою очередь, проблематика оптимизации вычислительной сложности порождения на базе исходно заданного множества эвристических алгоритмов корректного алгоритма, реализуемого в рамках математической техники так называемых корректных алгебр над множеством некорректных алгоритмов [9-11], и сегодня имеет ряд открытых вопросов и ожидающих своего решения математических задач.

Наконец, при использовании в процессе ИАД техники машинного обучения на примерах и контрпримерах [7, 8] в части нижних оценок сложности вычислений также приходится сталкиваться с доказуемо трудно-решаемыми комбинаторными задачами. Детальному обоснованию этого утверждения мы и уделим основное внимание ниже.

В работе [16] был предложен вариант уточнения сформулированной выше схемы **ПС**, в основе которого – техника, позволяющая восстановить покрытие исходно заданного для машинного обучения и восстановления зависимостей из данных средствами ДСМ-метода множества примеров (и контрпримеров) формируемыми классами эквивалентности (техника восстановления соответствующих классов сходства исходно заданных объектов а также покрывающих их классов эквивалентности). Общая схема используемой здесь конструкции выглядит следующим образом:

1. Каждый представленный в исходных данных объект может быть охарактеризован набором признаков (например, определенным набором тех или иных последовательностей 0 и 1 в его «теле»).

2. Все множество таких признаков может быть собрано в (пополняемый и расширяемый) каталог – «алфавит» («атомарных») признаков. Каждый уже идентифицированный ранее объект может быть «закодирован» (представлен) определенным набором (множеством) таких признаков.

3. На множестве идентификационных признаков могут быть заданы отношения (т.е. описания исходно заданных объектов могут стать не множествами признаков, а, например, графами, цепочками и т.п.). Описания изучаемых объектов могут быть расширены также числовыми значениями тех или иных параметров (*множества признаков + отношения + числовые характеристики*).

4. В случае пустоты исходно заданного множества контрпримеров, к имеющимся коллекциям объектов (примерам изучаемых объектов) путем дополнительного, - проводимого с привлечением экспертов по соответствующей проблематике, - анализа изучаемой предметной области могут быть добавлены также наборы контрпримеров – комплексов признаков из исходного «алфавита», которые, тем не менее, не являются объектами, обладающими целевыми признаками.

5. На множествах примеров и контрпримеров может быть определена (алгебраическая) операция сходства $*$, удовлетворяющая трем условиям:

- $a*a = a$
- $a*b = b*a$
- $(a*b)*c = a*(b*c) = a*b*c$

Например: $*$ определяемая как пересечение множеств, или же как выделение множества всех максимальных общих подграфов двух графов, множества всех максимальных общих подцепочек двух цепочек и т.п.

6. С помощью такой операции $*$ на примерах и контрпримерах строятся классы сходства: для каждого заданного объекта исходной коллекции примеров (и, соответственно, контрпримеров) класс сходства - это множество всех известных объектов, которые сходны с ним (т.е. операция сходства в каждом таком случае дает непустой - «ненулевой» - результат).

7. Диагностика (идентификация) новых объектов производится с помощью порождаемых таким образом классов сходства: возможность отнесения нового – диагностируемого -

объекта к какому-либо из уже реконструированных (по примерам или же контрпримерам) классов сходства позволяет утверждать, что это – еще один объект, аналогичный (см. класс сходства) ранее уже наблюдавшимся (в исходной обучающей выборке примеров и контрпримеров).

8. Более детально идея техники отнесения нового объекта к классу описывается следующим образом:

- класс сходства есть объединение некоторых классов эквивалентности (порождаемых всеми возможными в данном случае версиями соответствующего отношения эквивалентности) [17],
- по построенному классу сходства можно восстановить (с учетом имеющегося контекста – т.е. множества примеров и контрпримеров, задействованных в обучении) порождающие его классы эквивалентности;
- каждый класс сходства при восстановлении классов эквивалентности, объединением которых он является, будет характеризоваться непустыми (ненулевыми) результатами вычисления уже заданной операции сходства. Таким образом, каждый восстанавливаемый здесь класс эквивалентности будет определяться теми примерами из исходной выборки, результат вычисления операции сходства для которых будет одним и тем же² (например, в случае простого сходства на множествах – всех примеров, содержащих общее подмножество образующих); следовательно
- каждый класс эквивалентности в таком случае – это (в смысле [7, 8] и др.) так называемая орбита (множество примеров-родителей) для соответствующего сходства - ДСМ-гипотезы. (При этом, разумеется, один и тот же пример из исходной ДСМ-выборки может попадать в несколько пересекающихся, но не совпадающих классов эквивалентности для, разумеется, различных – см. различные ДСМ-гипотезы – конкретных вариантов отношений эквивалентности,

² Именно этим фактом дополнительно к рефлексивности (идемпотентности соответствующей операции сходства) и симметричности обеспечивается также и транзитивность соответствующего (уже используемой операции сходства) отношения сходства (суженного представленным дополнительным условием до эквивалентности).

участвующих в формировании соответствующих классов сходства);

- в дополнение к этому порождаемые на исходном множестве примеров и контрпримеров классы сходства, а также соответствующие классы эквивалентности (подробности описания структуры ДСМ-рассуждений, например, в работах [7, 8]) могут характеризоваться выполнимостью дополнительных логических условий – в частности: запрета на контрпримеры, условия единственности причины и т.п.

Теперь предметом нашего основного интереса будет построение оценок вычислительной сложности ряда задач порождения (по заданному множеству примеров) обсуждаемых ДСМ-классов эквивалентности с заданными характеристиками, а также задачи о вычислении мощности множества всех таких классов. Механизм построения соответствующих оценок вычислительной сложности мы подробно рассмотрим на примере работы с исходными объектами, представленными как множества «признаков» - образующих из некоторого наперед заданного «алфавита». Операция сходства в данном случае будет формализована как операция пересечения множеств $\cap$.

1. Исходные понятия и определения

Пусть $U = \{a_1, a_2, \dots, a_n\}$ универсум образующих (при этом, не теряя общности, мы будем считать, что в U нет одинаковых), а $\Omega = \{A_1, A_2, \dots, A_m\} \subseteq 2^U \setminus \emptyset$ - множество объектов (т.е. непустых множеств образующих), построенных над универсумом U (т.е. множество объектов, построенных из образующих $a_1, a_2, \dots, a_n$).

Определение 1

Множество $A_{i1,i2,\dots,is} = A_{i1} \cap A_{i2} \cap \dots \cap A_{is}$ мы будем называть *исчерпывающим сходством* объектов из множества Ω тогда и только тогда, когда:

$$A_{i1,i2,\dots,is} \neq \emptyset, \quad (1)$$

$$(\forall j) [(A_j \in \Omega) \& \neg(A_j \in \{A_{i1}, A_{i2}, \dots, A_{is}\})] \supset [(A_j \cap A_{i1,i2,\dots,is}) \subset A_{i1,i2,\dots,is}]^3 \quad (2)$$

³ Другими словами, каждое исчерпывающее сходство $A_{i1,i2,\dots,is}$ - не пусто и образовано в точности всеми теми объектами из Ω , которые при пересечении (вычислении сходства) дают именно множество $A_{i1,i2,\dots,is}$.

Пусть $A_{i1,i2,\dots,is} = A_{i1} \cap A_{i2} \cap \dots \cap A_{is} = \{a_{i1}, a_{i2}, \dots, a_{i1}\}$ - исчерпывающее сходство (пересечение). Мы будем называть объекты $A_{i1}, A_{i2}, \dots, A_{is}$ - *родителями* исчерпывающего сходства $A_{i1,i2,\dots,is}$, множество $\{A_{i1}, A_{i2}, \dots, A_{is}\}$ - *орбитой* исчерпывающего сходства $A_{i1,i2,\dots,is}$, а образующие $a_{i1}, a_{i2}, \dots, a_{i1}$ - *образующими* исчерпывающего сходства $A_{i1,i2,\dots,is}$. Множество всех исчерпывающих сходств, порождаемых над универсумом U из множества объектов Ω , мы будем обозначать посредством $ИС(\Omega)$.

Определение 2

Длиной исчерпывающего сходства будем называть мощность множества его образующих. *Размером орбиты* исчерпывающего сходства будем называть мощность множества его родителей.

Пусть $U = \{a_1, a_2, \dots, a_n\}$ - универсум образующих, а

$$\Omega = \{A_1, A_2, \dots, A_m\} \subseteq 2^U \setminus \emptyset$$

есть множество объектов (т.е. непустых множеств образующих), построенных над универсумом U (т.е. множество объектов, построенных из образующих $a_1, a_2, \dots, a_n$). Рассмотрим универсум $U^* = \{a^*_1, a^*_2, \dots, a^*_m\}$ и множество объектов

$$\Omega^* = \{A^*_1, A^*_2, \dots, A^*_n\} \subseteq 2^{U^*} \setminus \emptyset,$$

построенных над универсумом U^* по следующим правилам:

- пусть $A_i = \{a_{i1}, a_{i2}, \dots, a_{is}\} \subseteq U$ - i -тый элемент множества Ω : $A_i \in \Omega$.
- Каждому объекту A_i из множества Ω сопоставляется образующая a^*_i из универсума U^* , а каждому объекту A^*_i из множества Ω^* - образующая a_i из универсума U .
- Каждый объект A^*_i в множестве Ω^* сформирован теми и только теми образующими a^*_i из множества U^* , для которых сопоставленные им объекты A_i содержат сопоставленную объекту A^*_i образующую a_i .

Определение 3

Пусть заданы универсум U и Ω - некоторое множество объектов над ним. Построенные по описанным выше правилам множества U^* и Ω^* будем называть (соответственно) сопряженным универсумом и сопряженным множеством объектов над ним.

Рассмотрим некоторые свойства сопряженных множеств объектов над сопряженными универсумами. Очевидным следствием определения сопряженности универсумов и соответствующих множеств объектов над ними является

Лемма 1

Пусть заданы U и Ω . Тогда:

$$[U^*]^* = U \text{ и } [\Omega^*]^* = \Omega.$$

Более сложные взаимоотношения сопряженных универсумов представляет

Теорема 1 (о двойственности)

Пусть заданы: универсум U , множество объектов над ним - Ω и натуральное число k . В множестве $ИС(\Omega)$ всех исчерпывающих сходств объектов из Ω существует исчерпывающее сходство длины k тогда и только тогда, когда в множестве $ИС(\Omega^*)$ всех исчерпывающих сходств объектов из Ω^* найдется исчерпывающее сходство с орбитой размера k .

Доказательство

$\Rightarrow$. Предположим противное. Тогда в $ИС(\Omega)$ существует некоторый элемент A_0 вида $A_0 = \{a_{i1}, a_{i2}, \dots, a_{ik}\}$, но при этом в $ИС(\Omega^*)$ не существует элемента A'_0 вида

$$A'_0 = \{A^*_{i1} \cap A^*_{i2} \cap \dots \cap A^*_{ik}\}.$$

Рассмотрим множество A^*_0 , представляющее собой пересечение объектов $A^*_{i1}, A^*_{i2}, \dots, A^*_{ik}$ из Ω^* . Т.к.

$$A_0 \neq \emptyset, \text{ то } A^*_{i1} \cap A^*_{i2} \cap \dots \cap A^*_{ik} = A^*_0 \neq \emptyset,$$

кроме того, т.к. в U нет повторяющихся элементов, то A^*_0 - исчерпывающее сходство, а размер его орбиты - ровно k элементов. Это противоречит сделанному нами предположению (об отсутствии A'_0 заданного вида: достаточно положить $A'_0 = A^*_0$).

$\Leftarrow$. Предположим противное. Тогда в $ИС(\Omega^*)$ найдется элемент вида

$$A^*_0 = A^*_{i1} \cap A^*_{i2} \cap \dots \cap A^*_{ik},$$

но при этом в $ИС(\Omega)$ не должно оказаться элемента длины ровно k . Рассмотрим множество

$$A_0 = \{a_{i1}, a_{i2}, \dots, a_{ik}\}.$$

Для каждого j из множества $\{i1, i2, \dots, ik\}$ рассмотрим соответствующее подмножество Ω_j объектов из множества Ω , причем таких, что каждый из них содержит соответствующую об-

разующую a_j . Т.к. $A^*_0 \in ИС(\Omega^*)$, то каждое из таких подмножеств Ω_j не пусто, а их пересечение есть в точности A_0 . Это противоречит сделанному нами предположению.

Из Теоремы 1 и Леммы 1 очевидным образом вытекает

Следствие 1

Пусть заданы: универсум U , множество объектов над ним - Ω и натуральное число k . В множестве исчерпывающих сходств объектов из Ω существует исчерпывающее сходство с орбитой размера k тогда и только тогда, когда в множестве исчерпывающих сходств объектов из Ω^* найдется исчерпывающее сходство длины k .

Полезным для последующих рассмотрений утверждением является также

Лемма 2

Пусть заданы U и Ω . Пусть также $A_{i1, i2, \dots, is} = A_{i1} \cap A_{i2} \cap \dots \cap A_{is}$ есть произвольное исчерпывающее сходство объектов $A_{i1}, A_{i2}, \dots, A_{is}$ из Ω , формирующее множество родителей этого исчерпывающего сходства $A_{i1, i2, \dots, is}$. Тогда множество $\{A_{i1}, A_{i2}, \dots, A_{is}\}$ есть класс эквивалентности на Ω , сформированный по отношению E , определяемому как «иметь (содержать) общую часть $A_{i1, i2, \dots, is}$ ».

Доказательство

Рефлексивность и симметричность отношения E есть следствие наличия этих свойств у используемой нами операции сходства – пересечения множеств $\cap$. Транзитивность E есть следствие *исчерпываемости* сходства (см. выше - **Определение 1**): если $A_{i,j,k}$ есть общая часть объектов A_i и A_j , и одновременно $A_{i,j,k}$ есть общая часть объектов A_j и A_k , то $A_{i,j,k}$ есть общая часть также и объектов A_i и A_k .

Таким образом, оперируя с исчерпывающими сходствами, мы фактически имеем дело с двумя характеристиками каждого соответствующего класса эквивалентности $E(A_{i1, i2, \dots, is})$ (участвующего в покрытии порождаемых на множестве примеров Ω операцией $\cap$ классов сходства):

- *числом его элементов* (размером множества родителей соответствующего исчерпывающего сходства $A_{i1, i2, \dots, is}$) и

- числом образующих, которые формируют рассматриваемое (и, фактически, порождающее этот класс эквивалентности) исчерпывающее сходство $A_{i1,i2,\dots,is}$ – его длиной (см. выше - **Определение 2**).

Определим пару отображений s и g , связывающих множества U и Ω^4 следующим образом:

$$s : 2^\Omega \rightarrow 2^U$$

и для

$$\forall X = \{A_{i1}, A_{i2}, \dots, A_{is}\} \in 2^\Omega$$

имеем

$$s(X) = A_{i1} \cap A_{i2} \cap \dots \cap A_{is};$$

кроме того

$$g : 2^U \rightarrow 2^\Omega$$

и для

$$\forall Y = \{a_{i1}, a_{i2}, \dots, a_{ik}\} = \{A^*_{i1}, A^*_{i2}, \dots, A^*_{ik}\} \in 2^U$$

имеем

$$g(Y) = A^*_{i1} \cap A^*_{i2} \cap \dots \cap A^*_{ik}$$

Легко видеть, что отображения s и g удовлетворяют соотношениям:

$$\begin{aligned} [X_1 \subseteq X_2] &\supseteq [s(X_2) \subseteq s(X_1)] \\ [Y_1 \subseteq Y_2] &\supseteq [g(Y_2) \subseteq g(Y_1)], \end{aligned}$$

т.е. представляют собой соответствие Галуа [17, 18]. Несложно также убедиться, что отображения $s^*g(X) = g(s(X))$ и $g^*s(Y) = s(g(Y))$ являются операторами замыкания, соответственно, на множествах Ω и U . Теперь, наконец, не составляет труда проверить, что **Определение 1** выделяет замкнутые (в смысле операторов s^*g и g^*s) подмножества, соответственно в универсумах U^* (а вместе с ним - и в множестве Ω) и U . Итак, множество неподвижных точек замыкания Галуа s^*g (и, соответственно, g^*s) над множеством Ω (и, соответственно, над Ω^*) определяет множество всех исчерпывающих сходств объектов из Ω (и, соответственно, из Ω^*).

“Наглядное” представление о природе исчерпывающих сходств пересечений дает следующая комбинаторная интерпретация:

- пусть задан универсум образующих $U = \{a_1, a_2, \dots, a_n\}$ и множество объектов $\Omega = \{A_1, A_2, \dots, A_m\} \subseteq 2^U \setminus \emptyset$ над универсумом U ;
- рассмотрим n -мерный куб BC_n , поместив на нем m вершин, сопоставленных элементам множества Ω по следующему закону:

- каждому $A_i = \{a_{i1}, a_{i2}, \dots, a_{is}\} \in \Omega$ сопоставляется вершина из BC_n , характеризующаяся таким набором единиц и нулей, что единицы в ней стоят на позициях $i1, i2, \dots, is$ и только на них;

- если среди элементов множества Ω есть одинаковые, то им сопоставляется одна и та же вершина куба BC_n ;

• пересечению элементов любого подмножества мощности $m \geq k \geq 2$ в множестве Ω сопоставляются (и помечаются):

- вершина $\langle 0, 0, \dots, 0 \rangle$ в ситуации когда рассматриваемое пересечение пусто;

- вершина, порождаемая единицами на тех позициях, номера которых - номера образующих из U , попавших в вычисленное пересечение;

• на множестве помеченных вершин $DV(BC_n, \Omega)$ естественным образом (по вложению сопоставленных помеченных вершин множеств образующих из U) определяется отношение частичного порядка α ;

• теперь помеченный фрагмент n -мерного куба BC_n представляет диаграмму частично упорядоченного множества $\langle DV(BC_n, \Omega), \alpha \rangle$, в которой (непустому) исчерпывающему сходству объектов $A_{i1}, A_{i2}, \dots, A_{is}$ из множества Ω соответствует $A_{i1,i2,\dots,is} = A_{i1} \cap A_{i2} \cap \dots \cap A_{is}$ - точная нижняя граница (в смысле частичного порядка α) множества вершин n -мерного куба BC_n , сопоставленных объектам $A_{i1}, A_{i2}, \dots, A_{is}$, выделенная вместе со всеми A_0 - максимальными элементами частично упорядоченного множества $\langle DV(BC_n, \Omega), \alpha \rangle$, для которых выполняется вложимость:

$$\emptyset \subset A_{i1,i2,\dots,is} \subseteq A_0.$$

2. Представления о числе ДСМ-сходств, порождаемых из заданного множества примеров

Пусть заданы универсум U и множество объектов над ним - Ω . Обратимся к более подробному анализу свойств множества $GC(U, \Omega)$ исчерпывающих сходств (неподвижных точек определенных нами замыканий Галуа) для заданных множеств U и Ω . Предварительные представления о возможном “объеме” множества $GC(U, \Omega)$ дает

⁴ Более точно: связывающих множества U, U^*, Ω и Ω^* .

Лемма 3

Существуют U_0 и Ω_0 такие, что мощность множества $GC(U_0, \Omega_0)$ растет экспоненциально быстро с ростом размеров множеств U_0 и Ω_0 .

Доказательство

Пусть $U_0 = \{a_1, a_2, \dots, a_n\}$, а $\Omega_0 = \{A_1, A_2, \dots, A_n\}$, где

$$A_i = U_0 \setminus \{a_i\} = \{a_1, a_2, \dots, a_n\} \setminus \{a_i\}.$$

Легко видеть, что сходство (пересечение) любого непустого подмножества объектов из Ω представляет собой соответствующее (т.е. порожденное именно этими объектами и только ими) исчерпывающее сходство. Число таких исчерпывающих сходств

$$S = |GC(U_0, \Omega_0)| = \sum_{i=1}^{n-2} C^n_i = 2^n - n - 2 = O(2^n)$$

(здесь перечислены исчерпывающие сходства, порожденные каждым i образующими из универсума U_0).

Определение 4

Задачей **ЧИСЛО КЛАССОВ ЭКВИВАЛЕНТНОСТИ** (или **ЧИСЛО КЭ**) будем называть следующую задачу:

Дано: множества U и Ω .

Найти: число элементов в множестве $GC(U, \Omega)$

$$S = |GC(U, \Omega)|$$

Точное представление о комбинаторной природе множества $GC(U, \Omega)$ дает

Теорема 2

Задача **ЧИСЛО КЭ** пересчитательно полна:

$$\text{ЧИСЛО КЭ} \in \#PC$$

(здесь $\#PC$ представляет собой класс пересчитательно полных - например, [19-22] и др. - комбинаторных задач).

Доказательство

Идея доказательства - продемонстрировать полиномиальную сводимость известной [21] переборной задачи **МОНОТОННАЯ ВЫПОЛНИМОСТЬ** (понимаемой как задача вида:

Дано: монотонная булевская функция Φ в виде 2-КНФ

$$\Phi(x_1, x_2, \dots, x_n) = \Phi_1 \& \Phi_2 \& \dots \& \Phi_m = (x_{11} \vee x_{12}) \& (x_{21} \vee x_{22}) \& \dots \& (x_{m1} \vee x_{m2})$$

Найти: число наборов значений переменных $x_1, x_2, \dots, x_n$, выполняющих функцию Φ)

к рассматриваемой нами задаче **ЧИСЛО КЭ**. Мы реализуем следующую схему сведения:

- по заданной функции Φ полиномиально быстро строится пара множеств U_Φ, Ω_Φ , такая, что исчерпывающие пересечения объектов из Ω_Φ над универсумом U_Φ “перечисляют” все нули функции Φ ;

- размер множества единиц функции Φ (т.е. множества наборов, выполняющих эту функцию) есть размер дополнения множества нулей функции Φ до множества всех возможных наборов значений переменных $x_1, x_2, \dots, x_n$.

Теперь более подробно. Пусть Φ имеет вид

$$\Phi(x_1, x_2, \dots, x_n) = \Phi_1 \& \Phi_2 \& \dots \& \Phi_m = (x_{11} \vee x_{12}) \& (x_{21} \vee x_{22}) \& \dots \& (x_{m1} \vee x_{m2}).$$

Не теряя общности, будем считать, что в представлении функции Φ все конъюнкты Φ_i различны. (В противном случае функция Φ может быть приведена к нужному нам виду полиномиально сложной процедурой - сортировкой образующих ее конъюнктов)

Положим

$$U_\Phi = \{a_1, a_2, \dots, a_n\}, \text{ а } \Omega_\Phi = \{A_1, A_2, \dots, A_m, A_{m+1}\},$$

при чем так, что

$$A_j = \{a_1, a_2, \dots, a_n\} \setminus \{a_{i1}, a_{i2}\} \text{ для } i \in \{1, 2, \dots, n\} \text{ и } j \in \{1, 2, \dots, m\},$$

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

Фактически, каждый из объектов A_i (для $1 \leq i \leq m$) сопоставляется нулю соответствующего конъюкта Φ_i из 2-КНФ представления функции Φ .

В ситуации, когда функция Φ обращается в нуль, по крайней мере, один из ее конъюнктов непременно также обращается в нуль. Будем называть каждую ситуацию, когда ровно один из образующих Φ конъюнктов обращается в нуль, “примитивным” нулем функции Φ . Нетрудно убедиться, что каждый из нулей функции Φ может быть рассмотрен как суперпозиция “примитивных” нулей (т.е. ситуация, когда одновременно реализуются соответствующие “примитивные” нули функции Φ).

Легко видеть, что сходства (пересечения) вида

$$A_i \cap A_{m+1}$$

являются исчерпывающими (в функции Φ нет повторяющихся конъюнктов !) и могут быть сопоставлены “примитивным” нулям функции Φ . Каждое из пересечений более чем двух объектов из множества Ω_Φ также будет представлять какой-либо из нулей функции Φ . И, наоборот, каждый из нулей функции Φ - как соответствующая суперпозиция “примитивных” нулей функции Φ - может быть представлен (исчерпывающим - !, ведь в Φ нет повторяющихся конъюнктов) сходством объектов из множества Ω_Φ . Таким образом, множество нулей функции Φ и множество исчерпывающих сходств объектов из множества Ω_Φ могут быть взаимно однозначным образом сопоставлены друг другу. Т.е. зная число исчерпывающих сходств

$$S_\Phi = |GC(U_\Phi, \Omega_\Phi)|,$$

мы имеем возможность полиномиально быстро определить число наборов, выполняющих функцию Φ .

Для завершения доказательства не составляет труда убедиться, что множества U_Φ и Ω_Φ полиномиально быстро порождаются по заданной функции Φ .

3. Основные переборные задачи поиска ДСМ-сходств

Экспоненциально быстрый (с ростом U и Ω) рост размеров множества $GC(U, \Omega)$ заставляет обратиться к анализу достаточно просто обозримых частных характеристик множества исчерпывающих пересечений, порождаемых над заданным универсумом U из множества примеров Ω . Одной из таких характеристик естественно считать границы множества $GC(U, \Omega)$ - т.е. множества максимальных и минимальных по вложению элементов в $GC(U, \Omega)$.

Определение 5

1. Задача **ВСЕ ИСЧЕРПЫВАЮЩИЕ СХОДСТВА МАКСИМАЛЬНОГО РАЗМЕРА** определяется следующими условиями:

Дано: множества U и Ω .

Найти: множество всех максимальных по вложению элементов в $GC(U, \Omega)$.

2. Задача **ВСЕ ИСЧЕРПЫВАЮЩИЕ СХОДСТВА МИНИМАЛЬНОГО РАЗМЕРА** определяется следующими условиями:

Дано: множества U и Ω .

Найти: множество всех минимальных по вложению элементов в $GC(U, \Omega)$.

3. Задача **ВСЕ КЛАССЫ ЭКВИВАЛЕНТНОСТИ МАКСИМАЛЬНОГО РАЗМЕРА** определяется следующими условиями:

Дано: множества U и Ω .

Найти: множество всех элементов в $GC(U, \Omega)$, характеризуемых максимальным по вложению размером порождаемых ими классов эквивалентности.

4. Задача **ВСЕ КЛАССЫ ЭКВИВАЛЕНТНОСТИ МИНИМАЛЬНОГО РАЗМЕРА** определяется следующими условиями:

Дано: множества U и Ω .

Найти: множество всех элементов в $GC(U, \Omega)$, характеризуемых минимальным по вложению размером порождаемых ими классов эквивалентности.

5. Задача **ИСЧЕРПЫВАЮЩЕЕ СХОДСТВО РАЗМЕРА НЕ МЕНЕЕ k** определяется следующими условиями:

Дано: натуральное число k , множества U и Ω .

Найти: элемент множества $GC(U, \Omega)$, содержащий не менее k образующих из множества U .

6. Задача **ИСЧЕРПЫВАЮЩЕЕ СХОДСТВО РАЗМЕРА НЕ БОЛЕЕ k** определяется следующими условиями:

Дано: натуральное число k , множества U и Ω .

Найти: элемент множества $GC(U, \Omega)$, содержащий не более k образующих из множества U .

7. Задача **КЛАСС ЭКВИВАЛЕНТНОСТИ РАЗМЕРА НЕ МЕНЕЕ k** определяется следующими условиями:

Дано: натуральное число k , множества U и Ω .

Найти: элемент множества $GC(U, \Omega)$, которому соответствует класс эквивалентности размера не менее k (элементов из множества Ω).

8. Задача **КЛАСС ЭКВИВАЛЕНТНОСТИ РАЗМЕРА НЕ БОЛЕЕ k** определяется следующими условиями:

Дано: натуральное число k , множества U и Ω .

Найти: элемент множества $GC(U, \Omega)$, которому соответствует класс эквивалентности размера не более k (элементов из множества Ω).

9. Задача **ИСЧЕРПЫВАЮЩЕЕ СХОДСТВО РАЗМЕРА РОВНО k** определяется следующими условиями:

Дано: натуральное число k , множества U и Ω .

Найти: элемент множества $GC(U, \Omega)$, содержащий ровно k образующих из множества U .

10. Задача **КЛАСС ЭКВИВАЛЕНТНОСТИ РАЗМЕРА РОВНО k** определяется следующими условиями:

Дано: натуральное число k , множества U и Ω .

Найти: элемент множества $GC(U, \Omega)$, которому соответствует класс эквивалентности размера ровно k (элементов из множества Ω).

4. Оценки сложности вычислений для теоретико-множественного случая

Вычислительную сложность порождения верхней границы каждого из названных выше множеств исчерпывающих сходств (а с ними - и соответствующих классов эквивалентности) характеризует

Теорема 3

Задача **ВСЕ ИСЧЕРПЫВАЮЩИЕ СХОДСТВА МАКСИМАЛЬНОГО РАЗМЕРА** полиномиально разрешима (т.е. существует алгоритм, порождающий ее решение в ходе вычислений, объем которых ограничен сверху полиномом от характерных размеров входных данных).

Доказательство

Легко видеть, что каждое исчерпывающее сходство максимального размера порождается не менее чем парой объектов из множества Ω . Таким образом, для порождения множества всех исчерпывающих сходств максимального размера достаточно построить соответствующие замыкания Галуа для всех пар объектов из множества Ω , удалив затем из построенного множества те замыкания Галуа, которые

вкладываются по крайней мере в один из других (отличных от него самого) элементов построенного множества.

Вычислительную сложность порождения нижней границы каждого множества исчерпывающих сходств (а с ними - и соответствующих классов эквивалентности) характеризует

Теорема 4

Задача **ВСЕ ИСЧЕРПЫВАЮЩИЕ СХОДСТВА МИНИМАЛЬНОГО РАЗМЕРА** полиномиально разрешима.

Доказательство

Легко видеть, что каждое исчерпывающее сходство минимального размера содержит по крайней мере одну образующую из универсума U . Таким образом, для порождения множества всех исчерпывающих сходств минимального размера достаточно для каждой образующей из универсума U найти сходство (пересечение) всех объектов из Ω , содержащих данную образующую, удалив затем из полученного множества те элементы, в которые вкладывается по крайней мере один из других (оставшихся) элементов этого же (полученного) множества - см. соображения о так называемых наклеенных образующих в работах [23-25].

Принимая во внимание Теорему 1 (о двойственности), легко убедиться, что Теорема 3 и Теорема 4 имеют

Следствие 2

Задача **ВСЕ КЛАССЫ ЭКВИВАЛЕНТНОСТИ МАКСИМАЛЬНОГО РАЗМЕРА** полиномиально разрешима.

Следствие 3

Задача **ВСЕ КЛАССЫ ЭКВИВАЛЕНТНОСТИ МИНИМАЛЬНОГО РАЗМЕРА** полиномиально разрешима.

Кроме этого из Теорем 3,4 и Следствий 2,3 вытекают также Следствия 4-7:

Следствие 4

Задача **ИСЧЕРПЫВАЮЩЕЕ СХОДСТВО РАЗМЕРА НЕ МЕНЕЕ k** полиномиально разрешима.

Доказательство

Достаточно рассмотреть элементы верхней границы множества $GC(U, \Omega)$, остановив рас-

смотрение на первом, удовлетворяющем необходимому условию. Нетрудно понять, что в случае отсутствия такового среди элементов верхней границы, его нет и во всем множестве $GC(U, \Omega)$.

Следствие 5

Задача **ИСЧЕРПЫВАЮЩЕЕ СХОДСТВО РАЗМЕРА НЕ БОЛЕЕ k** полиномиально разрешима.

Доказательство

Достаточно рассмотреть элементы нижней границы множества $GC(U, \Omega)$, остановив рассмотрение на первом, удовлетворяющем необходимому условию. Нетрудно понять, что в случае отсутствия такового среди элементов нижней границы, его нет и во всем множестве $GC(U, \Omega)$.

Следствие 6

Задача **КЛАСС ЭКВИВАЛЕНТНОСТИ РАЗМЕРА НЕ МЕНЕЕ k** полиномиально разрешима.

Доказательство

Искомый результат - следствие Теоремы 1 и Следствия 4.

Следствие 7

Задача **КЛАСС ЭКВИВАЛЕНТНОСТИ РАЗМЕРА НЕ БОЛЕЕ k** полиномиально разрешима.

Доказательство

Искомый результат - следствие Теоремы 1 и Следствия 5.

Комбинаторную природу задачи **ИСЧЕРПЫВАЮЩЕЕ СХОДСТВО РАЗМЕРА РАВНО k** характеризует

Теорема 5

Задача **ИСЧЕРПЫВАЮЩЕЕ СХОДСТВО РАЗМЕРА РАВНО k** принадлежит классу NP-полных задач.

Доказательство

Пусть задана произвольная булевская функция

$$\begin{aligned} \Phi(x_1, x_2, \dots, x_n) &= \Phi_1 \& \Phi_2 \& \dots \& \Phi_m = \\ &= (x^{\alpha_{11}}_{11} \vee x^{\alpha_{12}}_{12} \vee \dots \vee x^{\alpha_{1n}}_{1n}) \& (x^{\alpha_{21}}_{21} \vee x^{\alpha_{22}}_{22} \\ &\quad \vee \dots \vee x^{\alpha_{2n}}_{2n}) \& \dots \\ &\quad \dots \& (x^{\alpha_{m1}}_{m1} \vee x^{\alpha_{m2}}_{m2} \vee \dots \vee x^{\alpha_{mn}}_{mn}), \end{aligned}$$

где $x^{\alpha_i}_i = x_i$ при $\alpha_i = 1$, и $x^{\alpha_i}_i = \neg x_i$ при $\alpha_i = 0$.

Наша цель - построить универсум U_Φ и множество объектов над ним Ω_Φ так, чтобы

- отображение $f: \Phi \rightarrow \langle U_\Phi, \Omega_\Phi \rangle$ позволяло эффективно (т.е. с полиномиальными затратами ресурсов - размеров памяти и объема вычислений) порождать множества U_Φ и Ω_Φ по заданной функции Φ ;
- для функции Φ выполняющий ее набор значений булевских переменных существовал тогда и только тогда, когда в множестве $GC(U_\Phi, \Omega_\Phi)$ существовал бы элемент, порожденный ровно k образующими из множества U_Φ .

Для каждой входящей в конъюнктивную нормальную форму функции Φ переменной x_i (и ее отрицания $\neg x_i$) перечислим все те конъюнкты Φ_i из $\Phi(x_1, x_2, \dots, x_n)$, которые выполняются при обращении x_i (и соответственно $\neg x_i$) в единицу. Обозначим множество конъюнктов, выполняемых при выполнении литеры x_i , через $Q(x_i)$, а множество конъюнктов, выполняемых при выполнении литеры $\neg x_i$, через $q(\neg x_i)$. Рассмотрим множества

$$Q^*(x_i) = \{\Phi_1, \Phi_2, \dots, \Phi_m\} \setminus \{Q(x_i)\}$$

и

$$q^*(\neg x_i) = \{\Phi_1, \Phi_2, \dots, \Phi_m\} \setminus \{q(\neg x_i)\}.$$

Для построения искомого отображения f , прежде всего воспользуемся следующим очевидным обстоятельством:

если набор $\langle \alpha_1, \alpha_2, \dots, \alpha_n \rangle$ значений переменных $\langle x_1, x_2, \dots, x_n \rangle$ выполняет функцию Φ , то

$$\bigcup_{i=1}^n Z_i(x^{\alpha_i}_i) = \{\Phi_1, \Phi_2, \dots, \Phi_m\}$$

и, соответственно,

$$\bigcap_{i=1}^n Z_i^*(x^{\alpha_i}_i) = \emptyset$$

для

$$Z_i = Q \text{ при } \alpha_i = 1, \text{ и } Z_i = q \text{ при } \alpha_i = 0.$$

Пусть заданы множества

$$A = \{A_1, A_2, \dots, A_r\} \text{ и } a = \{a_1, a_2, \dots, a_r\}.$$

Определение 6

Множество $B \subset (A \cup a)$ будем называть ИС-допустимым, тогда и только тогда, когда

- 1) $|B| = r$
- 2) $\neg \exists j ((1 \leq j \leq r) \& (A_j \in B) \& (a_j \in B))$

Другими словами, ИС-допустимое подмножество объединения двух равномоощных конечных множеств (элементы которых пронумерованы) мощности r содержит ровно r элементов, среди которых нет ни одной пары элементов, имеющих один и тот же номер.

Несложно убедиться, что тот факт, что набор

$$\langle \alpha_1, \alpha_2, \dots, \alpha_n \rangle, \text{ где } \alpha_i \in \{0,1\} \text{ а } i \in \{1, \dots, n\}$$

выполняет функцию Φ , означает, что объединение элементов соответствующего ИС-допустимого множества, построенного из соответствующих Q_i и q_i совпадает с множеством $\{\Phi_1, \Phi_2, \dots, \Phi_m\}$. Более точно: выполнимость Φ означает, что объединение элементов соответствующего ИС-допустимого множества $B = \{b_1, b_2, \dots, b_n\}$, полученного из множеств

$$A = \{Q_1, Q_2, \dots, Q_n\} \text{ и } a = \{q_1, q_2, \dots, q_n\}$$

по правилу

$$b_i = \begin{cases} Q_i, & \text{при } \alpha_i = 1 \\ q_i, & \text{при } \alpha_i = 0. \end{cases} \quad (Пр)$$

дает множество

$$\{\Phi_1, \Phi_2, \dots, \Phi_m\}.$$

Таким образом, справедлива

Лемма 4

Пусть $B(\underline{\alpha}) = \{b_1, b_2, \dots, b_n\}$ - ИС-допустимое множество, полученное по правилу (Пр) из

$$A = \{Q_1, Q_2, \dots, Q_n\} \text{ и } a = \{q_1, q_2, \dots, q_n\}$$

по набору $\underline{\alpha} = \langle \alpha_1, \alpha_2, \dots, \alpha_n \rangle$ (где $\alpha_i \in \{0,1\}$ а $i \in \{1, \dots, n\}$), выполняющему функцию Φ . Тогда

$$\bigcup_{i=1}^n b_i = \{\Phi_1, \Phi_2, \dots, \Phi_m\}.$$

Нетрудно убедиться, что из Леммы 4 с учетом соотношения

$$\setminus(A \cup B) = \setminus A \cap \setminus B,$$

где $\setminus X$ есть дополнение множества X , вытекает

Лемма 5

Пусть $B^*(\underline{\alpha}) = \{b^*_1, b^*_2, \dots, b^*_n\}$ - ИС-допустимое множество, полученное по правилу (Пр*)

$$b^*_i = \begin{cases} Q^*_i, & \text{при } \alpha_i = 1 \\ q^*_i, & \text{при } \alpha_i = 0. \end{cases} \quad (Пр^*)$$

из

$$A^* = \{Q^*_1, Q^*_2, \dots, Q^*_n\} \text{ и } a^* = \{q^*_1, q^*_2, \dots, q^*_n\}$$

по набору $\underline{\alpha} = \langle \alpha_1, \alpha_2, \dots, \alpha_n \rangle$ (где $\alpha_i \in \{0,1\}$ а $i \in \{1, \dots, n\}$), выполняющему функцию Φ . Тогда

$$\bigcap_{i=1}^n b^*_i = \emptyset.$$

Рассмотрим множество Γ , образованное n парами множеств T_i, t_i ($i \in \{1, \dots, n\}$)

$$\Gamma = \{T_1, t_1, T_2, t_2, \dots, T_n, t_n\},$$

где

$$\begin{aligned} T_1 &= \{\Delta^1_1, \delta^1_1, \Delta^1_2, \Delta^2_2, \dots, \Delta^{n+1}_2, \dots \\ &\quad \Delta^1_n, \Delta^2_n, \dots, \Delta^{n+1}_n\} \\ t_1 &= \{\delta^1_2, \Delta^2_1, \Delta^1_2, \Delta^2_2, \dots, \Delta^{n+1}_2, \dots \\ &\quad \Delta^1_n, \Delta^2_n, \dots, \Delta^{n+1}_n\} \\ T_2 &= \{\Delta^1_1, \Delta^2_1, \dots, \Delta^{n+1}_1, \Delta^1_2, \delta^2_1, \dots \\ &\quad \Delta^1_n, \Delta^2_n, \dots, \Delta^{n+1}_n\} \\ t_2 &= \{\Delta^1_1, \Delta^2_1, \dots, \Delta^{n+1}_1, \delta^2_2, \Delta^2_2, \dots \\ &\quad \Delta^1_n, \Delta^2_n, \dots, \Delta^{n+1}_n\} \\ &\vdots \\ T_n &= \{\Delta^1_1, \Delta^2_1, \dots, \Delta^{n+1}_1, \Delta^1_2, \Delta^2_2, \dots, \Delta^{n+1}_2, \dots \\ &\quad \Delta^1_n, \delta^n_1\} \\ t_n &= \{\Delta^1_1, \Delta^2_1, \dots, \Delta^{n+1}_1, \Delta^1_2, \Delta^2_2, \dots, \Delta^{n+1}_2, \dots \\ &\quad \delta^n_2, \Delta^2_n\} \end{aligned}$$

Здесь в каждое множество T_i элемент δ^i_1 входит лишь в позиции номер $(i-1)(n+1)+2$, а элемент δ^i_2 множества t_i - лишь в позиции номер $(i-1)(n+1)+1$. Все элементы, имеющие различные индексы, различны.

Лемма 6

Пусть $\iota = \{t_{i1}, t_{i2}, \dots, t_{ir}\}$ - некоторое множество элементов из Γ . Пусть также

$$\varsigma = \bigcap_{j=1}^r t_{ij}$$

Множество ς содержит ровно n элементов тогда и только тогда, когда ι есть некоторое ИС-допустимое множество, построенное из множеств

$$T = \{T_1, T_2, \dots, T_n\} \text{ и } t = \{t_1, t_2, \dots, t_n\}$$

Доказательство

Пусть число элементов в множестве ς есть n . Покажем, что $r = n$ и в множестве ι нет T_j и t_j с одинаковым индексом j ($1 \leq j \leq n$).

а) Предположим, что $r < n$.

Пусть r_1 - число индексов j , таких, что T_j и t_j входят одновременно в $\mathbf{t}$. Тогда число элементов в множестве ζ есть

$$S \geq (n+1)(n-r) + (r-2r_1) + r_1(n+1) = \\ = (n+1)(n-r) + r + r_1(n-1) \geq n+1.$$

(Здесь первое слагаемое отвечает за элементы множества $\mathbf{f}$, которые не вошли в множество $\mathbf{t}$; второе - за вошедшие в $\mathbf{t}$ элементы, которым соответствуют различные индексы j ; третье - за элементы с одинаковым индексом j). Таким образом, мы имеем противоречие с исходным предположением, что $S=n$.

б) Предположим, что $r = n$, но число “повторов” $r_1 \neq 0$.

В этом случае

$$S \geq (n-2r_1) + r_1(n+1) = n + r_1(n-1) \geq n+1.$$

(Здесь первое слагаемое отвечает за элементы из множества $\mathbf{t}$, второе - за остальные, которых будет ровно r_1 штук). Таким образом, мы вновь имеем противоречие с исходным предположением, что $S=n$.

в) Предположим, что $r > n$.

в1) задействованы все n индексов j из множества $\{1, \dots, n\}$.

В этом случае имеется $r-n$ “повторов” в индексах, т.е.

$$S = n - (r-n) = 2n - r < n-1.$$

Таким образом, мы вновь имеем противоречие с исходным предположением, что $S = n$.

в2) задействованы r_1 из всех n индексов.

В этом случае (аналогично случаю **а**) имеем:

$$S = (n+1)(n-r_1) + r_1 - (r-r_1) \geq n+1 \text{ (ввиду того, что } r_1 - (r-r_1) \geq 0).$$

(Здесь первый член отвечает за те элементы множества $\mathbf{f}$, которые не вошли в множество $\mathbf{t}$; второй - за те, что вошли; третий - за повторы индексов среди элементов множества $\mathbf{t}$). Вновь имеем противоречие с исходным предположением, что $S=n$.

г). Предположим, что $r=n$ и среди индексов j у элементов множества $\mathbf{t}$ нет повторов.

В этом случае каждое из n штук δ^p_g ($p \in \{1, 2, \dots, n\}$, $g \in \{1, 2\}$) в пересечении по соответствующей позиции в множествах из $\mathbf{t}$ даст ровно $\emptyset$, т.е. из $2n$ “кандидатов на вхождение” в множество ζ останется только n , т.е.

$$S = |\zeta| = n.$$

Если $\mathbf{t}$ - ИС-допустимое множество, то аналогично случаю **г** из предыдущей части (т.е. $\Rightarrow$ -доказательства) множество ζ будет содержать ровно n символов.

Вернемся к рассмотрению множеств Q^*_i и q^*_i ($i \in \{1, \dots, n\}$). Каждому Q^*_i и q^*_i сопоставим, соответственно, множества

$$\{T^1_i, T^2_i, \dots, T^{m+1}_i\} \text{ и } \{t^1_i, t^2_i, \dots, t^{m+1}_i\},$$

полагая, что смена верхнего индекса у T и t означает замену соответствующих Δ и δ на новые, но по способу, описанному выше.

Рассмотрим множество $\omega = \{Y_1, Y_2, \dots, Y_{2n}\}$, где

$$Y_1 = Q^*_1 \cup T^1_1 \cup T^2_1 \cup \dots \cup T^{m+1}_1$$

$$Y_2 = q^*_1 \cup t^1_1 \cup t^2_1 \cup \dots \cup t^{m+1}_1$$

$$Y_3 = Q^*_2 \cup T^1_2 \cup T^2_2 \cup \dots \cup T^{m+1}_2$$

$$Y_4 = q^*_2 \cup t^1_2 \cup t^2_2 \cup \dots \cup t^{m+1}_2$$

.

.

.

$$Y_{2n-1} = Q^*_n \cup T^1_n \cup T^2_n \cup \dots \cup T^{m+1}_n$$

$$Y_{2n} = q^*_n \cup t^1_n \cup t^2_n \cup \dots \cup t^{m+1}_n$$

Пусть

$$w = \bigcup_{i=1}^{2n} Y_i.$$

Следующая цель наших построений - продемонстрировать взаимосвязь исчерпывающих сходств (пересечений) элементов множества ω и наборов $\underline{\alpha} = \langle \alpha_1, \alpha_2, \dots, \alpha_n \rangle$, где $\alpha \in \{0, 1\}$, $i \in \{1, \dots, n\}$, выполняющих функцию Φ .

Рассмотрим задачу **ИСЧЕРПЫВАЮЩЕЕ СХОДСТВО РАЗМЕРА РОВНО k - $\mathcal{E}_{\text{ис}}(w, \omega, k)$** . Пусть отображение

$$\varphi : \mathcal{E}_{\text{вып}}(m, n) \rightarrow \mathcal{E}_{\text{ис}}(w, \omega, k),$$

по каждой задаче **ВЫПОЛНИМОСТЬ БУЛЕВСКОЙ ФУНКЦИИ $\mathcal{E}_{\text{вып}}(m, n)$** - о выполнимости булевой функции

$$\Phi(x_1, x_2, \dots, x_n) = \Phi_1 \& \Phi_2 \& \dots \& \Phi_m = \\ = (x^{\alpha_{11}}_{11} \vee x^{\alpha_{12}}_{12} \vee \dots \vee x^{\alpha_{1n}}_{1n}) \& (x^{\alpha_{21}}_{21} \vee x^{\alpha_{22}}_{22} \\ \vee \dots \vee x^{\alpha_{2n}}_{2n}) \& \dots \\ \dots \& (x^{\alpha_{m1}}_{m1} \vee x^{\alpha_{m2}}_{m2} \vee \dots \vee x^{\alpha_{mn}}_{mn}),$$

где $x^{\alpha_i}_i = x_i$ при $\alpha_i = 1$, и $x^{\alpha_i}_i = \neg x_i$ при $\alpha_i = 0$ с параметрами m и n описанным выше (см. представленную выше последовательность порождения множеств $Y_1, Y_2, \dots, Y_{2n}$) способом стро-

ит задачу $\mathcal{E}_{\text{ис}}(w, \omega, k)$ о существовании среди исчерпывающих сходств элементов множества ω такого исчерпывающего сходства, в которое входили бы ровно $k = n(m+1)$ элементов из множества w .

Лемма 7

Порождение

$$\mathcal{E}_{\text{ис}}(w, \omega, k) = \mathcal{E}_{\text{ис}}[w(m, n), \omega(m, n), k(m, n)] = \mathcal{E}_{\text{ис}}(U_{\Phi}, \Omega_{\Phi}, k)$$

по Φ посредством φ требует объема вычислений, ограниченного сверху полиномом от размеров задачи $\mathcal{E}_{\text{вып}}(m, n)$.

Доказательство

Утверждение **Леммы** следует из того, что

- число элементов, необходимых для задания множеств $T^1_i, t^1_i, T^2_i, t^2_i, \dots, T^n_i, t^n_i$ не превосходит $2n+2n=4n$;
- число элементов в каждом из множеств Q^*_i и q^*_i (где $i \in \{1, \dots, n\}$) не превосходит m ;
- наконец, $|w| = m + 4n(m+1)$.

Лемма 8

Каждому решению $I_{\text{вып}}$ задачи $\mathcal{E}_{\text{вып}}(m, n)$ соответствует единственное решение $I_{\text{ис}}$ задачи $\mathcal{E}_{\text{ис}}(w, \omega, k) = \mathcal{E}_{\text{ис}}[w(m, n), \omega(m, n), k(m, n)] = \mathcal{E}_{\text{ис}}(U_{\Phi}, \Omega_{\Phi}, k)$ и наоборот.

Доказательство

Набор $\alpha = \langle \alpha_1, \alpha_2, \dots, \alpha_n \rangle$ выполняющий функцию Φ , однозначно определяет ИС-допустимое множество, построенное из Q^*_i и q^*_i (где $i \in \{1, \dots, n\}$), которое в свою очередь, однозначно определяет исчерпывающее сходство размера ровно $k(m, n) = n(m+1)$, получаемое из элементов множества $\omega(\Phi(m, n))$. (Действительно, в случае непустоты пересечения Y_j ($1 \leq j \leq s, s \leq n$) по элементам " Q^*_i - q^*_i -части", число элементов в получаемом аналогично множеству ζ (из **Леммы 6**) пересечении ζ^* не может быть кратным $m+1$, а значит и равным $n(m+1)$). Исчерпывающее сходство - множество ζ^* , полученное из элементов множества ω , однозначным образом определяет ИС-допустимое множество, состоящее из Q^*_i и q^*_i (где $i \in \{1, \dots, n\}$), которое, в свою очередь, однозначно определяет набор

$\alpha = \langle \alpha_1(Q^*_1, q^*_1), \alpha_2(Q^*_2, q^*_2), \dots, \alpha_n(Q^*_n, q^*_n) \rangle$, выполняющий функцию Φ :

$$\alpha_i = \begin{cases} 1, & \text{если в ИС-допустимое множество взято } Q^*_i \\ 0 & \text{если в ИС-допустимое множество взято } q^*_i. \end{cases}$$

Теперь очевидным образом справедливо

Следствие 8

- Множества решений задач $\mathcal{E}_{\text{вып}}(m, n)$ и

$$\begin{aligned} \mathcal{E}_{\text{ис}}(w, \omega, k=n(m+1)) &= \\ = \mathcal{E}_{\text{ис}}[w(m, n), \omega(m, n), k(m, n)=n(m+1)] &= \\ = \mathcal{E}_{\text{ис}}(U_{\Phi}, \Omega_{\Phi}, k) \end{aligned}$$

равномощны.

- Отображение φ консервативно в смысле [19] и [20].

Этим доказательство **Теоремы 5** завершено.

Продemonстрированное выше (в доказательстве **Теоремы 5**) взаимно однозначное соответствие (мы только что убедились, что используемая сводимость

$$f: \Phi \rightarrow \langle U_{\Phi}, \Omega_{\Phi} \rangle$$

носит консервативный - смотри, например, [19, 20] и др. - характер) элементов множества выполняющих функцию Φ наборов значений булевских переменных и множества исчерпывающих пересечений размера ровно k в $GC(U_{\Phi}, \Omega_{\Phi})$ доказывает, что имеет место

Следствие 9

Задача **ИСЧЕРПЫВАЮЩЕЕ СХОДСТВО РАЗМЕРА РОВНО k** принадлежит классу #PC - перечислительно полных задач.

В свою очередь, комбинаторную природу задачи **КЛАСС ЭКВИВАЛЕНТНОСТИ РАЗМЕРА РОВНО k** характеризует

Следствие 10

- Задача **КЛАСС ЭКВИВАЛЕНТНОСТИ РАЗМЕРА РОВНО k** принадлежит классу NPC (NP-полных задач).

- Задача **КЛАСС ЭКВИВАЛЕНТНОСТИ РАЗМЕРА РОВНО k** принадлежит классу #PC - перечислительно полных задач.

Доказательство

Искомый результат - следствие **Теоремы 1** и **Следствий 8-9**.

Принимая во внимание установленную Следствиями 9-10 перечислительную полноту задач **ИСЧЕРПЫВАЮЩЕЕ СХОДСТВО РАЗМЕРА РОВНО k** и **КЛАСС ЭКВИВАЛЕНТНОСТИ РАЗМЕРА РОВНО k** , мы имеем возможность убедиться также в перечислительной полноте и соответствующего множества переборных задач из **Определение 5**, получаемых заменой равенства $...=k$ на неравенства $... \geq k$ и $... \leq k$. (Действительно, при каждой из таких замен множество решений получаемой новой переборной задачи содержит в себе множество решений задачи для равенства $...=k$, принадлежащей, как только что было показано, NPC - классу NP-полных задач). Таким образом, имеет место

Следствие 11

Задачи

- **ИСЧЕРПЫВАЮЩЕЕ СХОДСТВО РАЗМЕРА НЕ МЕНЕЕ k**
- **ИСЧЕРПЫВАЮЩЕЕ СХОДСТВО РАЗМЕРА НЕ БОЛЕЕ k**
- **КЛАСС ЭКВИВАЛЕНТНОСТИ РАЗМЕРА НЕ МЕНЕЕ k**
- **КЛАСС ЭКВИВАЛЕНТНОСТИ РАЗМЕРА НЕ БОЛЕЕ k**

принадлежат классу #PC перечислительно полных задач.

Итак, здесь мы можем говорить об аналогии со сложностными характеристиками известных задачам о выполнимости произвольной булевой функции и о выполнимости булевой функции в виде монотонной 2-КНФ. В случае названных переборных задач переход от NP - полноты первой задачи к ее полиномиально разрешимому частному случаю - второй задаче - не оказывает влияния на принадлежность обеих рассматриваемых задач к классу перечислительно полных [21]. В рассмотренном нами только что выше случае подобным же образом ослабление равенства $...=k$ до любого из неравенств $... \geq k$ или $... \leq k$ переводит соответствующую переборную задачу (**Теорема 5** и **Следствия 9-10**) из класса NP - полных в класс полиномиально разрешимых задач. Тем не менее, такое упрощение поиска какого-либо одного решения соответствующей рассматриваемой задачи не сказывается на экспоненциально большом (в общем случае) размере множества всех таких решений. Модифицированные (и,

вообще говоря, - упрощенные) варианты рассматриваемых переборных задач остаются по-прежнему перечислительно полными.

Особого внимания заслуживает консервативный характер предложенной в доказательстве **Теоремы 5** сводимости задачи **ВЫПОЛНИМОСТЬ БУЛЕВСКОЙ ФУНКЦИИ** к задаче **ИСЧЕРПЫВАЮЩЕЕ СХОДСТВО РАЗМЕРА РОВНО k** . Действительно, каждому набору значений булевских переменных, выполняющих функцию Φ , взаимно однозначным образом сопоставляется соответствующее исчерпывающее пересечение из $GC(U_\Phi, \Omega_\Phi)$. При этом на вид функции Φ , вообще говоря, не накладывается никаких дополнительных ограничений. Иными словами, процедура сводимости корректна и для частных случаев представлений функции Φ . Таким образом, рассмотрев конкретный частный случай, когда Φ представлена в виде 2-КНФ, мы “автоматически” убеждаемся, что имеет место (полностью совпадающее с полученным совершенно иным путем утверждением **Теоремы 5**)

Следствие 12

Задача **ЧИСЛО ИС** перечислительно полна:

$$\text{ЧИСЛО ИС} \in \#PC$$

В заключение обсудим свойства еще одного частного вида элементов во множестве $GC(U, \Omega)$ над заданными U и Ω - исчерпывающих пересечений, для которых зафиксирована сумма двух характеризующих их параметров - длины исчерпывающего пересечения и размера его орбиты.

Пусть $\mathcal{E}_{\text{ип}}(U, \Omega, k)$ - задача **ИСЧЕРПЫВАЮЩЕЕ СХОДСТВО РАЗМЕРА РОВНО k** . Рассмотрим двойственную ей задачу $\mathcal{E}_{\text{ип}}^*(U, \Omega, k) = \mathcal{E}_{\text{ип}}(U^*, \Omega^*, l)$. Задачу $\mathcal{E}_{\text{ип}}^0(U^0, \Omega^0, m)$ определим следующими условиями:

- 1) $U^0 = U \cup U^*$
- 2) $\Omega^0 = \Omega \times \Omega^*$
- 3) $m = k + l$.

Лемма 9

Задача $\mathcal{E}_{\text{ис}}^0(U^0, \Omega^0, m)$ строится по задачам $\mathcal{E}_{\text{ис}}(U, \Omega, k)$ и $\mathcal{E}_{\text{ис}}^*(U, \Omega, k)$ за полиномиальное от размеров задачи $\mathcal{E}_{\text{ис}}(U, \Omega, k)$ число операций.

Доказательство

Утверждение Леммы следует (с учетом **Леммы 1**, **Теоремы 1** и **Следствия 1**) из того, что

$$1) |U^0| = |U| + |\Omega|$$

2) Ω^0 строится по Ω и Ω^* не более чем за $|\Omega| \times |\Omega^*|$ шагов.

Сопоставим решения задач $\mathcal{E}_{ис}(U, \Omega, k)$ и $\mathcal{E}_{ис}^0(U^0, \Omega^0, m)$:

если I - решение задачи $\mathcal{E}_{ип}(U, \Omega, k)$, такое, что его размер есть k , то сопоставим ему решение I^0 задачи $\mathcal{E}_{ис}^0(U^0, \Omega^0, m)$, такое, что его размер есть $k+l$.

Лемма 10

Каждому решению задачи $\mathcal{E}_{ис}(U, \Omega, k)$ соответствует одно и только одно решение задачи $\mathcal{E}_{ис}^0(U^0, \Omega^0, m)$ и наоборот.

Доказательство

Утверждение рассматриваемой Леммы следует из необходимости иметь в каждом решении задачи $\mathcal{E}_{ис}^0(U^0, \Omega^0, m)$ решение одновременно задач $\mathcal{E}_{ис}(U, \Omega, k)$ и $\mathcal{E}_{ис}^*(U, \Omega, k)$, декартово произведение которых и задает решение задачи $\mathcal{E}_{ис}^0(U^0, \Omega^0, m)$.

Пусть ϕ - отображение, порождающее $\mathcal{E}_{ип}(U^0, \Omega^0, m)$ по $\mathcal{E}_{ип}(U, \Omega, k)$.

Лемма 11

Отображение ϕ - консервативно.

Доказательство

Утверждение Леммы следует из утверждений Лемм 9-10.

Теперь мы имеем все основания утверждать, что имеет место

Теорема 6

Задача $\mathcal{E}_{ис}^0(U^0, \Omega^0, m)$ лежит в классах NPC и #PC.

Доказательство

Утверждение Теоремы следует из Теоремы 5 с учетом Леммы 9.

Литература

1. Интеллектуальный анализ данных. – Википедия. - http://ru.wikipedia.org/wiki/Data_mining
2. Predictive Analytics and Data Mining. - <http://www.sas.com/technologies/analytics/datamining/index.html>
3. Большие данные. – Википедия. - Большие данные. - http://ru.wikipedia.org/wiki/Большие_данные.

Забейайло 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: mزابezhailo@arccn.ru

4. Вапник В.Н., Червоненкис А.Я. Теория распознавания образов. Статистические проблемы обучения – М.: Наука, 1974. - 416 С.
5. Вапник В. Н. Восстановление зависимостей по эмпирическим данным. М.: Наука, 1979. – 448 С.
6. Вапник В.Н. (ред.) Алгоритмы и программы восстановления зависимостей. - М.: Наука, 1984. - 816 С.
7. Финн В.К. Индуктивные методы Д.С. Милля в системах искусственного интеллекта // Искусственный интеллект и принятие решений. - 2010. – Часть I: №3, С.3 -21.
8. Финн В.К. Индуктивные методы Д.С. Милля в системах искусственного интеллекта // Искусственный интеллект и принятие решений. - 2010. – Часть II: №4, С. 14-40.
9. Журавлев Ю.И. Корректные алгебры над множествами некорректных (эвристических) алгоритмов. Часть I, Кибернетика. — 1977. — № 4. — С. 5-17.
10. Журавлев Ю.И. Корректные алгебры над множествами некорректных (эвристических) алгоритмов. Часть II, Кибернетика. — 1977. — № 6. — С. 21-27.
11. Журавлев Ю.И. Корректные алгебры над множествами некорректных (эвристических) алгоритмов. Часть III, Кибернетика. — 1978. — № 2. — С. 35-43.
12. Искусственная нейронная сеть. – Википедия. – Искусственная нейронная сеть. – https://ru.wikipedia.org/wiki/Искусственная_нейронная_сеть
13. Kohonen T. Self-Organizing Maps. - Springer-Verlag: Berlin - Heidelberg - New York, 2001. – 501 P.
14. Генетический алгоритм. - Википедия. – Генетический алгоритм. – https://ru.wikipedia.org/wiki/Генетический_алгоритм
15. Емельянов В.В., Курейчик В.В., Курейчик В.М. Теория и практика эволюционного моделирования. — М: Физматлит, 2003. — 432 С.
16. Забейайло М.И. К вопросу о достаточности оснований для принятия результатов интеллектуального анализа данных средствами ДСМ-метода. – НТИ, Сер.2. – 2015. №1, С.1-9.
17. Гусакова С.М., Финн В.К. Сходства и правдоподобный вывод // Известия АН СССР. Сер. Техническая Кибернетика -1987. - № 5. - С.42-63.
18. Кон П. Универсальная алгебра. - М.: Мир, 1968. – 359 С.
19. Гэри М., Джонсон Д.С. Вычислительные машины и трудно-решаемые задачи. - М.: Мир. - 1982. - 416 С.
20. Simon J. On the difference between one and many. - Lect. Not. Comp. Sci. -V.52 (1977). - Pp. 480-491.
21. Valiant L.G. The complexity of enumeration and reliability problems. - SIAM J.Comput. - 8 (1979), №1. - Pp. 410-421.
22. Valiant L.G. The complexity of computing the permanent. - Theoretical Computer Science. - 8 (1979). - Pp. 189-201.
23. Забейайло М.И. О некоторых возможностях управления перебором в ДСМ-методе // Искусственный интеллект и принятие решений. – 2014. – Часть I: № 1, С.95 -110.
24. Забейайло М.И. О некоторых возможностях управления перебором в ДСМ-методе // Искусственный интеллект и принятие решений. – 2014. – Часть II: № 3, С.3 - 21.
25. Забейайло М.И. Приближенный ДСМ-метода на примерах. – Научно-техническая информация. Сер.2. – 2014. – №10, С. 1-12.