6. *Бурделев А. В., Никонов В. Г., Лапиков И. И.* Распознавание параметров узла защиты информации, реализованного пороговой k-значной функцией // Труды СПИИРАН. 2016. № 46. С. 108–127.
7. *Анашкина Н. В., Шурупов А. Н.* Экспериментальное сравнение алгоритмов Балаша и имитации отжига в задаче решения систем линейных неравенств // Прикладная дискретная математика. Приложение. 2014. № 7. С. 151–153.
8. *Анашкина Н. В., Шурупов А. Н.* Применение алгоритмов локального поиска к решению систем псевдобулевых линейных неравенств // Прикладная дискретная математика. Приложение. 2015. № 8. С. 136–138.

УДК 519.7 DOI 10.17223/2226308X/12/62

АЛГОРИТМ «БЕЗОПАСНОЙ» ДЕКОМПОЗИЦИИ ФОРМАЛЬНОГО КОНТЕКСТА

Ч. М. Монгуш

Исследуется #P-полная задача нахождения всех формальных понятий заданного формального контекста. Предлагается алгоритм, который на практике позволяет решать данную задачу за полиномиальное время. Алгоритм основан на методе «безопасной» декомпозиции формального контекста на части, названные боксами. При «безопасной» декомпозиции формального контекста на боксы ни одно формальное понятие исходного контекста не теряется и не возникают новые формальные понятия. Процесс декомпозиции направлен на последовательное уменьшение размеров боксов формального контекста и реализуется итерационно. Установлены правила остановки процесса декомпозиции формального контекста на боксы, гарантирующие полиномиальное время его работы: задание порогового значения на плотность боксов и числа итераций разложения.

Ключевые слова: *формальный контекст, формальное понятие, декомпозиция формального контекста, алгоритм декомпозиции.*

Введение

Объектно-признаковая таблица — модель представления данных некоторой предметной области, в которой каждый столбец соответствует некоторому признаку, а каждая строка определяет признаковое описание отдельного объекта. Такая модель часто используется при решении прикладных задач в интеллектуальном анализе данных, в том числе в анализе формальных понятий.

Анализ формальныйх понятий (АФП) — прикладная ветвь алгебраической теории решёток Г. Биркгофа [1]. Основные идеи АФП были сформулированы в начале 80-х годов XX века в работах Р. Вилле и Б. Гантера и развиты в исследованиях С. О. Кузнецова, С. А. Объедкова, Д. И. Игнатова [2 – 4]. В рамках АФП объектно-признаковая таблица представляется формальным контекстом, отражающим наличие или отсутствие признаков, характерных для изучаемого множества объектов, и моделируется 0,1-матрицей. В АФП формальные понятия определяются с помощью соответствий Галуа и представляют собой пары множеств вида (объём, содержание).

В настоящее время применение методов АФП ограничивается высокой трудоёмкостью процесса нахождения всех формальных понятий заданного формального контекста. В данной работе предлагается алгоритм, который на практике позволяет разложить данную задачу на подзадачи (уменьшенные копии исходной задачи) за полиномиальное время.

1. Основные положения анализа формальных понятий

Приведём основные положения и обозначения АФП [2, 3]. Пусть для предметной области определены два непустых конечных множества G и M объектов и признаков (или свойств) и задано непустое отношение инцидентности $I \subseteq G \times M$. Данное отношение содержит информацию о выполнимости свойств из M на объектах из G, т. е. $(g, m) \in I$ означает, что объект g обладает признаком m, и наоборот — признак m присущ объекту g. Тройку $K = (G, M, I)$ принято называть формальным контекстом.

Будем полагать, что множества G и M линейно упорядочены (например, лексикографически). В этом случае формальный контекст $K = (G, M, I)$ однозначно задается 0,1-матрицей $T = (t_{ij})$: $t_{ij} = 0$ при $(g_i, m_j) \notin I$ и $t_{ij} = 1$ при $(g_i, m_j) \in I$ $(i = 1, 2, \ldots, |G|$; $j = 1, 2, \ldots, |M|)$.

Выберем в $K = (G, M, I)$ два произвольных подмножества $A \subseteq G$ и $B \subseteq M$ и определим для них отображения $(\cdot)'$ Галуа: $A' = \bigcap\limits_{g \in A} g'$, $B' = \bigcap\limits_{m \in B} m'$, где $g' = \{m \in M : (g, m) \in I\}$; $m' = \{g \in G : (g, m) \in I\}$. Согласно этому определению, множество A' — набор признаков, присущих объектам из A, а множество B' задаёт семейство объектов, обладающих признаками из B. Двойное применение $(\cdot)'$ определяет оператор замыкания $(\cdot)''$ на 2^M в алгебраическом смысле. Множество $(B)''$ можно трактовать как набор признаков, которые неизменно появляются в объектах формального контекста $K = (G, M, I)$ вместе с признаками из B, причём это множество является наибольшим по включению в пределах этого формального контекста. Если $B = B''$, то B называется замкнутым множеством относительно оператора $(\cdot)''$.

Пара множеств (A, B), таких, что $A \subseteq G$, $B \subseteq M$, $A' = B$ и $B' = A$, называется формальным понятием формального контекста $K = (G, M, I)$ с объёмом A и содержанием B. Далее для краткости в ряде случаев определение «формальный» перед словами «контекст» или «понятие» будем опускать. Пара множеств (A, B) является формальным понятием тогда и только тогда, когда $A = A''$ и $B = B''$. Очевидно, что всякое формальное понятие уникально в заданном контексте, т. е. отличается от других формальных понятий объёмом и/или содержанием. Если формальный контекст представлен 0,1-матрицей T, то при $A \neq \varnothing$ и $B \neq \varnothing$ формальному понятию (A, B) отвечает максимальная полная подматрица матрицы T.

Обозначим через FC множество всех формальных понятий формального контекста $K = (G, M, I)$. Пусть $(A_1, B_1), (A_2, B_2) \in FC$. Множество FC частично упорядочено отношением $(A_1, B_1) \sqsubseteq (A_2, B_2) \Leftrightarrow A_1 \subseteq A_2$. Отметим, что последнее эквивалентно условию $B_2 \subseteq B_1$. Каждое формальное понятие $(A, B) \in FC$ определяет для исследуемой предметной области совокупность однородных объектов A со своим специфичным набором признаков B. Если $(G, \varnothing) \in FC$, $(\varnothing, M) \in FC$, то формальные понятия $(G, \varnothing)$, $(\varnothing, M)$ называются тривиальными.

Определим на FC операции пересечения $\sqcap$ и объединения $\sqcup$ через одноимённые теоретико-множественные операции $\cap$ и $\cup$ следующим образом:

$$(A_1, B_1) \sqcap (A_2, B_2) = (A_1 \cap A_2, (A_1 \cap A_2)'), \quad (A_1, B_1) \sqcup (A_2, B_2) = ((B_1 \cap B_2)', B_1 \cap B_2).$$

Тогда $(FC, \sqsubseteq)$ образует полную решётку $L = (FC, \sqcap, \sqcup)$, называемую в АФП решёткой формальных понятий контекста $K = (G, M, I)$.

2. Постановка задачи и метод «безопасной» декомпозиции

В рамках АФП задача нахождения всех формальных понятий формулируется следующим образом. Задан формальный контекста $K = (G, M, I)$. Требуется для K най-

ти множество FC. На сегодняшний день для определения множества FC разработано много алгоритмов, в их числе NextClosure, Close-by-One, Norris [3, 4]. Время их выполнения в худшем случае составляет $O(|FC| \cdot |G|^2 \cdot |M|)$. Поскольку величина $|FC|$ может экспоненциально зависеть от $|G|$ и $|M|$, время выполнения данных алгоритмов также может быть экспоненциальным. Повысить производительность можно путём применения метода «безопасной» декомпозиции формального контекста на боксы [5].

Пусть $g \in G$ и $m \in M$ — произвольные элементы контекста $K = (G, M, I)$. Пары множеств (g'', g') и (m', m'') образуют формальные понятия, первое из которых назовём объектным, а второе — признаковым формальным понятием контекста $K = (G, M, I)$. Обозначим через $O = \{(g'', g') : g \in G\} \subseteq FC$ множество всех объектных формальных понятий и через $S = \{(m', m'') : m \in M\} \subseteq FC$ — множество всех признаковых формальных понятий контекста $K = (G, M, I)$.

Пара формальных понятий $(g'', g') \in O$, $(m', m'') \in S$ определяет бокс $\omega = (m', g', J)$ контекста $K = (G, M, I)$, если $(g'', g') \sqsubseteq (m', m'')$, что эквивалентно $g'' \subseteq m'$ (или $m'' \subseteq g'$). Про такой бокс будем говорить, что он образован элементами $g \in G$ и $m \in M$. Далее вместо $\omega = (m', g', J)$ будем кратко писать $\omega = (m', g')$ или (m', g').

Утверждение 1 [6]. Для всякого формального контекста $K = (G, M, I)$ и любых $(g'', g') \in O$, $(m', m'') \in S$ отношение порядка $(g'', g') \sqsubseteq (m', m'')$ выполняется тогда и только тогда, когда $(g, m) \in I$.

Утверждение 1 устанавливает оценку числа боксов, получаемых на каждой итерации разложения: число различных боксов, порождаемых всевозможными элементами контекста $K = (G, M, I)$, не превышает веса 0,1-матрицы T, т. е. величины $\|T\|$ — числа единичных элементов этой матрицы. Очевидно, что $1 \leqslant \|T\| \leqslant |G| \cdot |M|$.

Будем говорить, что формальное понятие $(A, B) \in FC$ вложено в бокс (m', g') контекста $K = (G, M, I)$, и записывать $(A, B) \preceq (m', g')$, если $A \subseteq m'$, $B \subseteq g'$. Всякий бокс (m', g') не является пустым, поскольку, согласно определению бокса, он всегда содержит формальные понятия $(g'', g') \in O$ и $(m', m'') \in S$.

Утверждение 2 [6]. Всякое нетривиальное формальное понятие (A, B) контекста $K = (G, M, I)$, которое вложено в бокс (m', g'), образованный элементами $g \in G$ и $m \in M$, содержит эти элементы и их замыкания, т. е. если $(A, B) \preceq (m', g')$, то $g \in A$ и $m \in B$; $g'' \subseteq A$ и $m'' \subseteq B$.

Согласно утверждению 2, пару (g'', m'') можно рассматривать в качестве типичного представителя не только бокса (m', g'), но и всех формальных понятий контекста $K = (G, M, I)$, вложенных в этот бокс. Переход от боксов к их типичным представителям в большинстве случаев уменьшает на практике время выполнения алгоритмов нахождения всех формальных понятий для заданного формального контекста. Соответствие между боксами и формальными понятиями контекста устанавливает следующая теорема.

Теорема 1 [6]. Для всякого формального контекста $K = (G, M, I)$, множества FC всех его формальных понятий и любой пары множеств (A, B), $\varnothing \neq A \subseteq G$, $\varnothing \neq B \subseteq M$, справедливо:

1) если $(A, B) \in FC$, то в K существует бокс $\omega = (m', g')$, $g \in G$ и $m \in M$, возможно, не единственный, в который это формальное понятие вложено;
2) если (A, B) — формальное понятие некоторого бокса $\omega = (m', g')$ формального контекста K, то оно также принадлежит FC.

Согласно теореме 1, разложение контекста $K = (G, M, I)$ на боксы является «безопасным» для любого формального понятия из FC [5]. Очевидно, что процесс разложения заданного контекста на боксы может быть организован итерационно, поскольку каждый выявленный на первой итерации бокс можно рассматривать в качестве исходного контекста и вновь подвергать декомпозиции. Определим сложность процесса разложения и правила его остановки. Пусть $|m'| \cdot |g'|$ — размер бокса (m', g'), а $\|(m', g')\|$ — число его единичных элементов. Плотностью бокса (m', g') назовём величину $\sigma(m', g') = \|(m', g')\|/(|m'| \cdot |g'|)$. Верны естественные границы $0 < \sigma(m', g') \leqslant 1$.

Утверждение 3 [6]. Всякий бокс (m', g') с плотностью $\sigma(m', g') = 1$ содержит ровно одно нетривиальное формальное понятие (A, B) контекста $K = (G, M, I)$, совпадающее с ним, т. е. $A = m'$ и $B = g'$.

Из утверждения 3 следует, что бокс (m', g') с плотностью 1 вырождается в нетривиальное формальное понятие и не подлежит дальнейшему разложению. Заметим, что время формирования одного бокса для формального контекста $K = (G, M, I)$ составляет $\mathrm{O}(|G| \cdot |M|)$. В целом, время, необходимое на однократное разложение этого контекста на боксы, в худшем случае составляет $\mathrm{O}\big(\sigma(G, M)\,|G|^2 \cdot |M|^2\big)$.

3. Алгоритм «безопасной» декомпозиции формального контекста на боксы

Алгоритм FindBoxes (алгоритм 1) реализует метод «безопасной» декомпозиции формального контекста на боксы. Теоретическим обоснованием этого алгоритма являются теорема 1 и утверждения 1–3, входными данными служат исходный контекст $K = (G, M, I)$ и целое положительное число k — число итераций. Результат работы алгоритма FindBoxes: Ω — множество боксов и H — множество типичных представителей боксов, входящих в Ω.

Алгоритм FindBoxes включает следующие основные процедуры: Boxes, Delete, SearchChains. Процедура Boxes разлагает заданный бокс ω, плотность которого отлична от 1, на более мелкие боксы и находит для них типичных представителей. Процедура Delete удаляет кратные боксы и боксы, совпадающие с исходным. Процедура SearchChains выявляет вложенные боксы, выполняет построение взаимно непересекающихся цепей частично упорядоченного множества боксов Ω_1 и находит для этих цепей максимальные элементы. Данная процедура позволяет уменьшать число боксов, получаемых на каждой итерации разложения.

Если число итераций процесса декомпозиции равно k, то разложение можно осуществить за время $\mathrm{O}\Big(|G|^{2k} \cdot |M|^{2k}\Big)$; при $k = 1$ алгоритм FindBoxes выполняется за время $\mathrm{O}\big(|G|^2 \cdot |M|^2\big)$. Для дополнительного ограничения числа частей, получаемых на каждой итерации, можно устанавливать пороговое значение на плотность боксов, подлежащих дальнейшему разложению. Это достигается заменой на шаге 8 алгоритма FindBoxes условия $\sigma(\omega) \neq 1$ условием $\sigma(\omega) < \sigma_0$, где σ_0 — пороговое значение плотности боксов, которые подлежит дальнейшему разложению.

Для оценки результативности алгоритма FindBoxes проведены вычислительные эксперименты при $k = 1$ и без задания ограничения на плотность боксов. Эксперименты проводились с помощью программы FCACorpus [7], осуществляющей нахождение всех формальных понятий. Использовались контексты, сгенерированные случайным образом. Для каждого контекста $K = (G, M, I)$ осуществлялось нахождение множества FC всех формальных понятий без разбиения на боксы и с итеративным разбиением на боксы. Анализировались два случая при проверке вложенности боксов: случай 1 — проверка производится без типичных представителей боксов, случай 2 —

Алгоритм 1. FindBoxes

Вход: исходный контекст $K = (G, M, I)$, k — количество итераций.
Выход: Ω — множество боксов, H — множество типичных представителей боксов из Ω.

$\Omega_1 := (G, M, I)$ // *множество боксов, подлежащих дальнейшему разложению*
$\Omega_2 := \varnothing$ // *множество боксов, не подлежащих дальнейшему разложению*
$H_1 := (G'', M'')$ // *множество типичных представителей боксов, входящих в* Ω_1
$H_2 := \varnothing$ // *множество типичных представителей боксов, входящих в* Ω_2
Пока $(k \neq 0 \;\&\; \Omega_1 \neq \varnothing)$
  $Q := \varnothing$, $V := \varnothing$.
  Для всех $\omega \in \Omega_1$
    Если $\sigma(\omega) \neq 1$, **то**
      Boxes(ω, X, Y); $Q := Q \cup X$; $R := R \cup Y$,
    иначе
      $\Omega_2 := \Omega_2 \cup \omega$; $H_2 := H_2 \cup H_1$.
  $W_1 := Q$; $H_1 := R$; Delete $(\Omega_1 \cup \Omega_2, H_1 \cup H_2)$.
  Если $\Omega_1 \neq \varnothing$ **то**
    SearchChains(Ω_1, H_1).
  $k := k - 1$.
$\Omega := \Omega_1 \cup \Omega_2$; $H := H_1 \cup H_2$.

с помощью типичных представителей. Результаты эксперимента приведены в таблице, где N — количество образованных боксов; $|FC|$ — число найденных формальных понятий; t — время выполнения программы. Эксперименты выполнялись на компьютере с процессором Intel Core i7-720QM Processor (6M Cache, 1,60 ГГц) и ОЗУ размером 4 Гбайт.

Оценка эффективности процесса декомпозиции формального контекста

Случаи	Характеристика исходного контекста				Результаты		
	$\lvert G\rvert$	$\lvert M\rvert$	$\lVert T\rVert$	$\sigma(G, M)$	N	$\lvert FC\rvert$	t, мс
Без разложения на боксы	100	20	1000	0,5	–	4962	145125
С разложением на боксы (случай 1)					883	4962	2878
С разложением на боксы (случай 2)					883	4962	**2200**
Без разложения на боксы	200	30	2940	0,49	–	10567	794520
С разложением на боксы (случай 1)					2895	10567	97906
С разложением на боксы (случай 2)					2895	10567	**90908**

Из таблицы видно, что значения $|FC|$ в случаях без разложения и с разложением на боксы полностью совпадают; число боксов, образованных при разложении контекста, не превышает величины $\|T\|$; алгоритм FindBoxes даёт значительный выигрыш по времени: время выполнения программы FCACorpus при разложении контекста на боксы уменьшается в несколько раз.

Заключение

Представленный алгоритм FindBoxes реализует метод «безопасной» декомпозиции формального контекста и на практике позволяет разложить задачу нахождения всех формальных понятий на подзадачи за полиномиальное время. Алгоритм также приме-

ним для ускорения существующих алгоритмов решения родственных задач, связанных с нахождением максимальных полных подматриц 0,1-матрицы.

ЛИТЕРАТУРА

1. *Биркгоф Г.* Теория решеток. М.: Наука, 1984. 568 с.
2. *Ganter B. and Wille R.* Formal Concept Analyses: Mathematical Foundations. Springer Science and Business Media, 2012. 314 p.
3. *Ganter B. and Obiedkov S. A.* Conceptual Exploration. Berlin; Heidelberg: Springer, 2016. 315 p.
4. *Kuznetsov S. O. and Obiedkov S. A.* Comparing performance of algorithms for generating concept lattices // J. Exper. Theor. Artificial Intelligence. 2002. V. 14. No. 2. P. 189–216.
5. *Mongush Ch. M. and Bykova V. V.* On decomposition of a binary context without losing formal concepts // J. Siberian Federal University. Mathematics and Physics. 2019. No. 3. P. 323–330.
6. *Быкова В. В., Монгуш Ч. М.* Декомпозиционный подход к исследованию формальных контекстов // Прикладная дискретная математика. 2019. №. 44. С. 111–124.
7. *Монгуш Ч. М., Быкова В. В.* Программа FCACorpus концептуального моделирования тувинских текстов методами анализа формальных понятий. Свид. о гос. регистрации программы для ЭВМ № 2018618907, выдано Федеральной службой по интеллектуальной собственности РФ, 2018.

УДК 519.7 DOI 10.17223/2226308X/12/63

ОБ ИСПОЛЬЗОВАНИИ ТЕХНОЛОГИЙ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПРОВЕРКИ СТАТИСТИЧЕСКИХ СВОЙСТВ СИММЕТРИЧНЫХ КРИПТОГРАФИЧЕСКИХ АЛГОРИТМОВ

А. А. Перов

Рассматривается применение технологий машинного обучения к задачам криптографии, в частности проведению статистического анализа блочных шифров. Изложена идея адаптации шифртекстов к алгоритму модели нейронной сети Inception V3. Приведены результаты экспериментов.

Ключевые слова: *криптография, машинное обучние, статистический анализ, раунд шифрования, итеративные блочные шифры.*

Введение

Традиционно статистический анализ проводится с помощью тестов, определяющих степень случайности выходной последовательности [1] с применением методов математической статистики. Исследования показывают, что современные итеративные алгоритмы шифрования обеспечивают удовлетворительные статистические свойства на меньшем, чем полное, числе раундов. Использование технологий машинного обучения для решения подобных задач является новым направлением. Для разработки методики проведения статистического анализа была использована нейронная сеть Inception v3, обычно применяющаяся для распознавания и классификации графических образов. Inception v3 является свёрточной нейронной сетью, состоящей из 17 слоёв, обученной на большом количестве изображений из базы ImageNet.

1. Преобразование шифртекстов

Для решения задачи по обучению нейронной сети для статистического анализа выполнено преобразование зашифрованных сообщений в формат графических изображе-