

УДК 519.7

## О ЛОГИЧЕСКОМ АНАЛИЗЕ ДАННЫХ С ЧАСТИЧНЫМИ ПОРЯДКАМИ В ЗАДАЧЕ КЛАССИФИКАЦИИ ПО ПРЕЦЕДЕНТАМ<sup>1)</sup>

© 2019 г. Е. В. Дюкова<sup>1,\*</sup>, Г. О. Масляков<sup>2,\*\*</sup>, П. А. Прокофьев<sup>3,\*\*\*</sup>

<sup>1)</sup> 119333 Москва, ул. Вавилова, 40, ВЦ ФИЦ ИУ РАН, Россия;

<sup>2)</sup> 119991 Москва, Ленинские горы, 1, МГУ им. М.В. Ломоносова, Россия;

<sup>3)</sup> 101000 Москва, Малый Харитоньевский переулок, 4, ИМАШ РАН, Россия)

\*e-mail: edjukova@mail.ru

\*\*e-mail: gleb-mas@mail.ru

\*\*\*e-mail: p\_prok@mail.ru

Поступила в редакцию 04.04.2019 г.  
Переработанный вариант 04.04.2019 г.  
Принята к публикации 15.05.2019 г.

Актуальность исследования обусловлена существованием прикладных задач машинного обучения, качественное решение которых невозможно в рамках классической постановки логического анализа данных. На основе обобщения базовых понятий предложена схема синтеза корректных логических процедур классификации по прецедентам, ориентированная на задание отношений частичных порядков на множествах значений признаков. Показано, что в общем случае при построении процедур классификации возникает необходимость рассматривать одну из центральных труднорешаемых дискретных задач, а именно, задачу дуализации над произведением частичных порядков. Дана матричная формулировка дуализации над произведением частичных порядков. Эффективность предлагаемого подхода к задаче классификации по прецедентам проиллюстрирована на модельных данных. Библ. 22. Фиг. 2.

**Ключевые слова:** логический анализ данных, классификация по прецедентам, монотонная дуализация, дуализация над произведением частичных порядков, неприводимое покрытие булевой матрицы, упорядоченное тупиковое покрытие целочисленной матрицы.

DOI: 10.1134/S0044466919090084

### ВВЕДЕНИЕ

В задаче классификации под прецедентной (обучающей) информацией понимается совокупность примеров изучаемых объектов, в которой каждый объект представлен в виде числового вектора, полученного на основе измерения или наблюдения ряда его параметров или характеристик. Подлежащие измерению или наблюдению свойства объектов называются признаками. В самом простом случае примеры делятся на два класса (класс положительных и класс отрицательных примеров). В общем случае число классов может быть больше двух. Требуется по признаковому описанию предъявленного объекта, о котором заранее неизвестно, какому классу он принадлежит, определить (распознать) этот класс.

Основное достоинство логического подхода к задаче классификации (распознавания) — возможность получения результата при отсутствии дополнительных предположений вероятностного характера и при небольшом числе прецедентов. Не требуется также задание метрики в пространстве описаний объектов. В данном случае для каждого признака определяется бинарная функция близости между его значениями, что позволяет проводить сравнение описания распознаваемого объекта с описаниями прецедентов. При этом большое внимание уделяется вопросам синтеза корректных алгоритмов, т.е. алгоритмов безошибочно классифицирующих материал обучения.

Модели тестовых алгоритмов распознавания, алгоритмов голосования по представительным наборам (алгоритмов типа “Кора”), алгоритмов голосования по покрытиям классов, классификаторов на основе решающих деревьев наиболее эффективны в случае целочисленной информа-

<sup>1)</sup> Работа выполнена при частичной финансовой поддержке РФФИ (проект 19-01-00430-а).

ции низкой значности, в особенности бинарной. Для их описания наравне с комбинаторным аппаратом применяется аппарат логических функций. Анализ прецедентной информации сводится к поиску в данных определенных закономерностей, различающих прецеденты из разных классов. Найденные закономерности (корректные элементарные классификаторы), как правило, имеют содержательное описание в терминах той прикладной области, в которой решается задача. По их наличию или, наоборот, отсутствию в описании распознаваемого объекта, решается вопрос о его классификации. Как правило, ищутся тупиковые (не избыточные) элементарные классификаторы. Первые модели рассматриваемых логических классификаторов предложены в [1]–[4]. Описание этих моделей с использованием понятия элементарного классификатора впервые приведено в [5], [6].

Существуют сложные задачи, когда не удается найти достаточное количество “информативных” корректных элементарных классификаторов. Подобная ситуация возникает, например, в случае целочисленных данных высокой значности. Один из способов решения проблемы – применение логических корректоров. Имеются в виду логические классификаторы, при построении которых используются методы так называемого алгебраического подхода [7]–[9].

Построение логических процедур классификации особенно актуально в тех случаях, когда набор большого числа прецедентов затруднен или вообще невозможен. Например, такая ситуация имеет место в задачах прогнозирования месторождений редких металлов, свойств твердых сплавов и т.д. Однако при больших размерах признакового пространства возникает необходимость рассматривать сложные в вычислительном плане задачи, которые в теории алгоритмической сложности дискретных задач называют труднорешаемыми задачами. Среди этих задач центральное место принадлежит монотонной дуализации – задаче построения максимальных конъюнкций монотонной булевой функции, заданной конъюнктивной нормальной формой. Задача допускает гиперграфовую формулировку [10] и матричную формулировку с использованием понятия неприводимого покрытия булевой матрицы [11]–[13]. Труднорешаемость монотонной дуализации имеет два аспекта: экспоненциальный рост числа решений при увеличении размера задачи и сложность их нахождения (перечисления). Наиболее эффективными считаются алгоритмы с полиномиальным шагом (алгоритмы с полиномиальной задержкой). Такие алгоритмы построены лишь для некоторых частных случаев монотонной дуализации (см., например, [14]). В общем случае лидерами по скорости счета являются асимптотически оптимальные алгоритмы, имеющие теоретическое обоснование, базирующиеся на получении асимптотических оценок типичного числа решений задачи и типичного числа “лишних” шагов алгоритма. Эти алгоритмы впервые предложены в [11].

Современные прикладные задачи классификации не всегда могут быть описаны в рамках классической постановки логической классификации, когда отдельные значения признака сравниваются с использованием простых отношений “равно”, “меньше”, “больше”. Предварительное изучение предметной области позволяет выявить более сложные отношения на множествах допустимых значений признаков. Довольно часто встречаются задачи классификации, в которых каждый признак принимает значения из некоторого частично упорядоченного множества.

В настоящей работе предложена схема синтеза корректных логических алгоритмов классификации при условии, что на множествах значений целочисленных признаков заданы частичные порядки. Обобщены базовые понятия, используемые при логическом анализе целочисленных данных в задаче классификации по прецедентам, и приведены условия корректности основных логических процедур классификации. Установлено, что анализ прецедентной информации с частичными порядками приводит к необходимости решать задачу дуализации над произведением конечных частичных порядков, простейшим частным случаем которой является монотонная дуализация. Дана матричная формулировка задачи дуализации общего вида и показано, что эта задача сводится к перечислению специальных покрытий целочисленной матрицы, названных в работе упорядоченными тупиковыми покрытиями. Понятие упорядоченного тупикового покрытия целочисленной матрицы – обобщение известного понятия неприводимого покрытия булевой матрицы, используемого в матричной формулировке задачи монотонной дуализации. На модельных данных показана зависимость качества логической классификации от выбора частичных порядков на множествах значений признаков.

## 1. ЛОГИЧЕСКАЯ КЛАССИФИКАЦИЯ ЦЕЛОЧИСЛЕННЫХ ДАННЫХ В КЛАССИЧЕСКОЙ ПОСТАНОВКЕ

В самых общих чертах задача классификации по прецедентам заключается в следующем. Исследуется некоторое множество объектов  $M$ . Известно, что  $M$  представимо в виде объединения  $l$  подмножеств  $K_1, \dots, K_l$ , называемых классами. Объекты из множества  $M$  описываются признаками  $x_1, \dots, x_n$ . Каждый признак имеет ограниченное число значений. Имеется конечный набор  $S_1, \dots, S_m$  объектов из множества  $M$ , о которых известно, каким классам они принадлежат. Это прецеденты или обучающие объекты. Пусть их описания имеют вид  $S_1 = (a_{11}, \dots, a_{1n})$ ,  $S_2 = (a_{21}, \dots, a_{2n})$ , ...,  $S_m = (a_{m1}, \dots, a_{mn})$ , здесь  $a_{ij}$  – значение признака  $x_j$  для объекта  $S_i$ . Требуется по предъявленному набору значений признаков  $(a_1, \dots, a_n)$ , описывающему некоторый объект  $S$  из  $M$ , о котором, вообще говоря, не известно, какому классу он принадлежит, определить этот класс.

Введем основные понятия, которые используются при построении классических логических процедур распознавания. Будем предполагать, что множество допустимых значений каждого признака состоит из целых чисел и это множество конечно.

Пусть  $H$  – набор из  $r$  различных признаков вида  $H = \{x_{j_1}, \dots, x_{j_r}\}$ ,  $\sigma = (\sigma_1, \dots, \sigma_r)$ ,  $\sigma_i$  – допустимое значение признака  $x_{j_i}$ ,  $i = 1, 2, \dots, r$ . Пара  $(\sigma, H)$  называется *элементарным классификатором* (эл.кл.) *ранга*  $r$ .

Близость объекта  $S = (a_1, \dots, a_n)$  из  $M$  и эл.кл.  $(\sigma, H)$ ,  $\sigma = (\sigma_1, \dots, \sigma_r)$ ,  $H = (x_{j_1}, \dots, x_{j_r})$ , оценивается величиной  $B(\sigma, S, H)$ , равной 1, если  $a_{j_i} = \sigma_i$  при  $i = 1, 2, \dots, r$ , и равной 0 в противном случае. Если  $B(\sigma, S, H) = 1$ , то говорят, что объект  $S$  порождает (содержит) эл.кл.  $(\sigma, H)$ .

Эл.кл.  $(\sigma, H)$  является *корректным для класса*  $K$ ,  $K \in \{K_1, \dots, K_l\}$ , если нельзя указать пару обучающих объектов  $S'$  и  $S''$  таких, что  $S' \in K$ ,  $S'' \notin K$  и  $B(\sigma, S', H) = B(\sigma, S'', H) = 1$ .

Таким образом, класс  $K$  не имеет корректных эл.кл., если существуют два прецедента  $S'$  и  $S''$  таких, что  $S' \in K$ ,  $S'' \notin K$  и описание  $S'$  совпадает с описанием  $S''$ . Поэтому предполагается, что любые два класса множества  $M$  не пересекаются.

Набор эл.кл.  $U$  называется *корректным для класса*  $K$ , если нельзя указать пару обучающих объектов  $S'$  и  $S''$  таких, что  $S' \in K$ ,  $S'' \notin K$  и  $B(\sigma, S', H) = B(\sigma, S'', H)$  для любого эл.кл.  $(\sigma, H)$  набора  $U$ .

Логические классификаторы можно условно разделить на два типа. Первый тип – это модели классификаторов, основанные на голосовании по корректным эл.кл. Второй тип – это классификаторы, основанные на голосовании по корректным наборам эл.кл. или логические корректоры.

### 1.1. Модели логических классификаторов, основанные на голосовании по корректным эл.кл.

Считается, что каждый распознающий алгоритм  $A$  на этапе обучения строит для каждого класса  $K$  некоторое множество корректных эл.кл.  $C^A(K)$ . Распознавание объекта  $S$  осуществляется на основе вычисления величины  $B(\sigma, S, H)$  для каждого построенного эл.кл.  $(\sigma, H)$ , т.е. каждый элемент множества  $C^A(K)$ ,  $K \in \{K_1, \dots, K_l\}$ , участвует в процедуре голосования. В результате для класса  $K$  вычисляется оценка  $\Gamma(S, K)$  принадлежности объекта  $S$  классу  $K$ . Таким образом, алгоритм  $A$  из рассматриваемого семейства распознающих алгоритмов определяется множествами  $C^A(K_1), \dots, C^A(K_l)$ . Алгоритмы отличаются и способом вычисления оценки  $\Gamma(S, K)$ . Рассмотрим основные модели.

В общем случае элементарный классификатор  $(\sigma, H)$ , по отношению к классу  $K$  может обладать одним из следующих трех свойств: 1) каждый обучающий объект из класса  $K$  содержит  $(\sigma, H)$ ; 2) не все, а лишь некоторые обучающие объекты из класса  $K$  содержат  $(\sigma, H)$ ; 3) ни один обучающий объект из класса  $K$  не содержит  $(\sigma, H)$ .

Первый случай встречается крайне редко, поэтому работать с эл.кл. первого типа не представляется возможным. Во втором и третьем случаях аргументом за отнесение распознаваемого объекта  $S$  в класс  $K$  считается ситуация, когда распознаваемый объект  $S$  соответственно содержит  $(\sigma, H)$  и не содержит  $(\sigma, H)$ .

Корректный эл.кл. второго типа называется *представительным эл.кл. класса*  $K$ . Представительный для класса  $K$  эл.кл.  $(\sigma, H)$  называется *тупиковым представительным эл.кл. класса*  $K$ , если

не является представительным для  $K$  любой эл.кл. вида  $(\sigma', H')$ , где  $\sigma' = (\sigma_1, \dots, \sigma_{t-1}, \sigma_{t+1}, \dots, \sigma_r)$ ,  $H' = H \setminus \{x_{j_t}\}$ ,  $t \in \{1, 2, \dots, r\}$ .

Понятие эл.кл., представительного эл.кл. и тупикового представительного эл.кл. может быть дано в терминах теории логических функций. Поясним сказанное на примере бинарных признаков.

В рассматриваемом случае эл.кл.  $(\sigma, H)$ ,  $H = \{x_{j_1}, \dots, x_{j_r}\}$ ,  $\sigma = (\sigma_1, \dots, \sigma_r)$ , — это элементарная конъюнкция (ЭК)  $B$  над переменными  $x_1, \dots, x_n$  вида  $x_{j_1}^{\sigma_1} \wedge \dots \wedge x_{j_r}^{\sigma_r}$ , которая обращается в 1 на описании объекта  $S$ , если объект  $S$  содержит эл.кл.  $(\sigma, H)$ , т.е. описание объекта  $S$  принадлежит интервалу истинности  $N_B$  конъюнкции  $B$ .

Пусть  $f_K(x_1, \dots, x_n)$  — частичная булева функция, определенная на описаниях объектов обучающей выборки и принимающая значение 1 только на описаниях обучающих объектов из  $K$ . Пусть  $N_{\bar{f}_K}$  — множество булевых наборов, на которых  $f_K$  принимает значение 0. ЭК  $B$  называется допустимой для  $f_K$ , если  $N_B \cap N_{\bar{f}_K} = \emptyset$ . ЭК  $B$  называется максимальной для функции  $f_K$ , если не существует допустимой для  $f_K$  ЭК  $B'$  такой, что  $N_{B'} \subset N_B$ .

Нетрудно видеть, что представительный эл.кл. класса  $K$  — это допустимая конъюнкция для функции  $f_K$ . Тупиковый представительный эл.кл. для класса  $K$  — это максимальная конъюнкция для  $f_K$ .

Набор признаков  $H$  называется тестом, если в каждом классе  $K$ ,  $K \in \{K_1, \dots, K_l\}$ , каждый прецедент содержит представительный для  $K$  эл.кл. вида  $(\sigma, H)$ . Тест называется тупиковым, если любое его собственное подмножество тестом не является.

Эл.кл. третьего типа называется покрытием класса  $K$ . Покрытие  $(\sigma, H)$  класса  $K$  называется тупиковым покрытием класса  $K$ , если не является покрытием для  $K$  любой эл.кл. вида  $(\sigma', H')$ , где  $\sigma' = (\sigma_1, \dots, \sigma_{t-1}, \sigma_{t+1}, \dots, \sigma_r)$ ,  $H' = H \setminus \{x_{j_t}\}$ ,  $t \in \{1, 2, \dots, r\}$ .

Переформулируем понятие представительного эл.кл. класса  $K$ , используя понятие покрытия класса. Пусть  $\bar{K} = M \setminus K$ . Будем рассматривать  $\bar{K}$  как отдельный класс, т.е. будем считать, что у нас всего два класса  $K$  и  $\bar{K}$ . Тогда, как нетрудно видеть, (тупиковый) представительный эл.кл. является (тупиковым) покрытием для класса  $\bar{K}$  и не является покрытием для класса  $K$ .

Модели тестовых алгоритмов и алгоритмов голосования по представительным эл.кл. основаны на построении для каждого класса  $K$  множества представительных эл.кл. этого класса  $C^A(K)$ . В простейших модификациях этих моделей число голосов, поданных эл.кл. из  $C^A(K)$  за принадлежность объекта  $S$  к классу  $K$ , вычисляется по формуле

$$\Gamma_1(S, K) = \frac{1}{|C^A(K)|} \sum_{(\sigma, H) \in C^A(K)} P_{(\sigma, H)} B(\sigma, S, H),$$

где  $P_{(\sigma, H)}$  — вес эл.кл.  $(\sigma, H)$ . В качестве  $P_{(\sigma, H)}$  обычно берется число обучающих объектов из  $K$ , содержащих  $(\sigma, H)$ . По похожей схеме работают полные корректные решающие деревья [15].

В моделях голосования по (тупиковым) покрытиям класса множество  $C^A(K)$  состоит из (тупиковых) покрытий класса  $K$ . Эл.кл. из  $C^A(K)$  голосует за принадлежность распознаваемого объекта классу  $K$ , если этот эл.кл. не встречается в описании распознаваемого объекта  $S$ . Принадлежность объекта  $S$  классу  $K$  в простейшей модификации оценивается величиной

$$\Gamma_2(S, K) = \frac{1}{|C^A(K)|} \sum_{(\sigma, H) \in C^A(K)} (1 - B(\sigma, S, H)).$$

Сравнение описанных выше моделей по качеству распознавания приведено в [5]. Наиболее информативными являются эл.кл. небольшого ранга. Поэтому при решении прикладных задач, как правило, либо ограничивают ранг эл.кл., либо рассматривают только тупиковые корректные эл.кл. (при этом не обязательно все). Например, алгоритм КОРА, предложенный в [2], использует эл.кл. с рангом меньшим или равным трем. Модель голосования по покрытиям класса выигрывает по скорости счета, если число прецедентов из  $K$  существенно меньше числа прецедентов из  $\bar{K}$ .

### 1.2. Логические корректоры

В задачах с большой значностью признаков, как правило, почти все корректные эл.кл. имеют большой ранг и, как следствие, каждый такой эл.кл. содержится в небольшом числе прецедентов. При этом под значностью признака понимается число его различных значений и вещественнозначные данные часто трактуются как целочисленные высокой значности. Задачи, в которых значность признаков слишком большая, сложны для классических логических алгоритмов распознавания. Существуют различные способы решения этой проблемы. Один из этих способов основан на идеях алгебраического подхода [16] и базируется на использовании произвольных эл.кл. (не обязательно корректных эл.кл.). Алгебраический подход применяется тогда, когда требуется скорректировать работу нескольких алгоритмов, каждый из которых безошибочно классифицирует лишь часть обучающих объектов. Цель коррекции – сделать так, чтобы ошибки одних алгоритмов были скомпенсированы другими алгоритмами. И качество результирующего алгоритма оказалось лучше, чем качество каждого из базовых алгоритмов в отдельности. Об алгебро-логическом подходе говорят, когда каждый базовый алгоритм однозначно определяется некоторым эл.кл. и корректирующие функции являются булевыми [7]–[9]. Оценки принадлежности распознаваемого объекта к классам вычисляются с использованием процедуры голосования по корректным наборам эл.кл.

Рассмотрим набор эл.кл.  $U = \{(\sigma_1, H_1), \dots, (\sigma_t, H_t)\}$ . Бинарный вектор  $U(S) = (q_1(S), \dots, q_t(S))$ , в котором  $q_i(S) = B(\sigma_i, S, H_i)$  при  $i \in \{1, 2, \dots, t\}$ , называется *откликом* набора  $U$  на объекте  $S$ . Очевидно, что набор эл.кл.  $U$  является корректным для класса  $K$  тогда и только тогда, когда не существует двух прецедентов с одинаковыми откликами, один из которых принадлежит  $K$ , а другой  $K$  не принадлежит. Очевидно также, что для корректного для класса  $K$  набора эл.кл.  $U$  всегда существует частичная булева функция  $F_{U,K}$ , выполняющая роль корректирующей. Функция  $F_{U,K}$  определена на откликах прецедентов и принимает значение 1 только на откликах прецедентов из  $K$ .

Особо следует отметить корректные наборы эл.кл., имеющие в качестве корректирующей функции монотонную булеву функцию. Такие наборы называются монотонными. Очевидным является следующее

**Утверждение 1.** *Набор эл.кл.  $U$  является монотонным корректным набором для класса  $K$  тогда и только тогда, когда для любых двух обучающих объектов  $S'$  и  $S''$ ,  $S' \in K$ ,  $S'' \notin K$ , в наборе  $U$  можно указать эл.кл.  $(\sigma, H)$  такой, что  $B(\sigma, S', H) = 1$ ,  $B(\sigma, S'', H) = 0$ .*

Заметим, что если набор эл.кл.  $U$  является монотонным корректным набором для класса  $K$  и состоит из одного эл.кл.  $(\sigma, H)$ , то  $(\sigma, H)$  – представительный эл.кл. для класса  $K$ . Набор признаков  $H$ , являющийся тестом, порождает для каждого класса  $K$  корректный набор эл.кл., в котором каждый эл.кл. – представительный эл.кл. для  $K$ , имеющий вид  $(\sigma, H)$ .

Логический корректор на этапе обучения для каждого класса  $K$  строит семейство  $W_K$ , состоящее из корректных для  $K$  наборов эл.кл. Далее осуществляется голосование по наборам из семейства  $W_K$ . Корректность классификации обеспечивается за счет корректности каждого голосующего набора эл.кл.

Распознавание объекта  $S$  по корректному набору эл.кл.  $U$  класса  $K$  осуществляется следующим образом. Для каждого объекта  $S'$  из обучающей выборки, принадлежащего классу  $K$ , выписывается вектор  $U(S')$ , который сравнивается с вектором  $U(S)$  (сравниваются отклики набора  $U$  на объектах  $S$  и  $S'$ ). Далее запись  $U(S') \preceq U(S)$  означает, что каждая координата вектора  $U(S')$  не превосходит соответствующую координату вектора  $U(S)$ . Объект  $S$  не получает голос за принадлежность к классу  $K$ , если не выполнено  $U(S') \preceq U(S)$ . Если  $U$  – монотонный корректный набор эл.кл. класса  $K$  и  $U(S') \preceq U(S)$ , то объект  $S$  получает голос за принадлежность к классу  $K$ . Если же  $U$  не является монотонным корректным набором эл.кл. класса  $K$ , то  $S$  получает голос за принадлежность к классу  $K$  только в случае  $U(S') = U(S)$ .

На практике проверено, что целесообразно использовать корректные наборы эл.кл. небольшой длины, в частности, тупиковые. Корректный набор эл.кл. для класса  $K$  называется *тупиковым*, если любое его собственное подмножество не является корректным для  $K$  набором эл.кл. Решение прикладных задач показывает, что монотонный логический корректор (логический корректор, использующий только монотонные корректные наборы эл.кл.) работает лучше немонотонного логического корректора.

На этапе построения семейств корректных наборов эл.кл. также, как и при построении логических процедур классификации, основанных на голосовании по корректным эл.кл., приходится решать сложные дискретные задачи, среди которых центральное место принадлежит задаче

монотонной дуализации. Как правило, используется матричная формулировка монотонной дуализации и поиск искоемых корректных эл.кл. или корректных наборов эл.кл. сводится к поиску неприводимых покрытий булевой матрицы, которая специальным образом строится по обучающей выборке. В [17] показано, что при поиске корректных эл.кл. можно обойтись без построения вспомогательной булевой матрицы, если ввести понятие (тупикового) покрытия целочисленной матрицы.

В простейших моделях логических корректоров для снижения вычислительной сложности делается следующее. Обучающая выборка делится на базовую и настроечную подвыборки. По базовой подвыборке строятся корректные наборы эл.кл., по настроечной подвыборке оценивается распознающая способность этих наборов. При этом для каждого класса с помощью генетического алгоритма отбираются наборы эл.кл. ранга 1 с распознающей способностью, близкой к максимальной.

В более сложных моделях используются эл.кл. произвольного ранга и для сокращения временных затрат строятся так называемые локальные базисы классов. Локальному базису соответствует корректный набор эл.кл. достаточно большой мощности, который используется для построения семейства тупиковых корректных наборов эл.кл., обладающих хорошей распознающей способностью. Локальные базисы либо строятся итеративно использованием бустинга, либо стохастическим способом.

В [9] предложена общая схема синтеза корректных логических процедур классификации, в рамках которой описаны алгоритмы голосования по корректным эл.кл. и существующие логические корректоры. В этой же работе построен и исследован логический корректор с поляризуемой булевой функцией в качестве корректирующей. Монотонная булева функция является частным случаем поляризуемой булевой функции.

## 2. ЛОГИЧЕСКИЙ АНАЛИЗ ДАННЫХ С ЧАСТИЧНЫМИ ПОРЯДКАМИ. ЗАДАЧИ ПОИСКА МАКСИМАЛЬНЫХ И МИНИМАЛЬНЫХ НЕЗАВИСИМЫХ ЭЛЕМЕНТОВ ПРОИЗВЕДЕНИЯ ЧАСТИЧНЫХ ПОРЯДКОВ

В данном разделе введены основные понятия логического анализа данных с частичными порядками, и сформулированы две центральные задачи такого анализа, а именно, задача поиска максимальных независимых элементов произведения частичных порядков и задача поиска минимальных независимых элементов произведения частичных порядков, простейшим частным случаем которой является задача монотонной дуализации.

Пусть  $P = P_1 \times \dots \times P_n$ , где  $P_1, \dots, P_n$  – конечные частично упорядоченные множества. Считается, что элемент  $y = (y_1, \dots, y_n) \in P$  следует за элементом  $x = (x_1, \dots, x_n) \in P$  ( $x$  предшествует  $y$ ), если  $y_i$  следует за  $x_i$  ( $x_i$  предшествует  $y_i$ ) при  $i = 1, 2, \dots, n$ . Для обозначения того, что  $y \in P$  следует за  $x \in P$  ( $x$  предшествует  $y$ ) далее используется запись  $x \preceq y$  или  $y \succeq x$ . Запись  $x < y$  ( $y > x$ ) означает, что  $x \preceq y$  и  $y \neq x$ . Элементы  $x, y \in P$  называются сравнимыми, если либо  $x \preceq y$ , либо  $y \preceq x$ . В противном случае  $x$  и  $y$  называются несравнимыми.

Пусть  $R \subseteq P$ . Введем обозначения:  $R^+ = R \cup \{x \in P \mid \exists a \in R, a < x\}$  – множество элементов, следующих за элементами из  $R$ ;  $R^- = R \cup \{x \in P \mid \exists a \in R, x < a\}$  – множество элементов, предшествующих элементам из  $R$ .

Элемент  $x$  множества  $P \setminus R^+$  ( $P \setminus R^-$ ) называется *максимальным (минимальным) независимым* от  $R$  элементом множества  $P$ , если для любого другого элемента  $y$  множества  $P \setminus R^+$  ( $P \setminus R^-$ ) отношение  $x < y$  ( $y < x$ ) не выполняется.

Обозначим через  $I(R^+)$  множество, состоящее из максимальных независимых от  $R$  элементов множества  $P$ , через  $I(R^-)$  – множество, состоящее из минимальных независимых от  $R$  элементов множества  $P$ .

Ставятся задачи построения для заданного  $R$  множеств  $I(R^+)$  и  $I(R^-)$ . Каждую из этих двух задач можно свести к другой путем изменения порядка на обратный.

Одним из наиболее востребованных и изученных является случай, когда множества  $P_1, \dots, P_n$  — цепи, т.е. в каждом из этих множеств любые два элемента сравнимы. Антицепью называется множество, в котором любые два различных элемента несравнимы. Множества  $I(R^+)$  и  $I(R^-)$  являются антицепями.

Простейшим случаем дуализации над произведением цепей является монотонная дуализация. Задача формулируется следующим образом.

Дана конъюнктивная нормальная форма (КНФ), реализующая монотонную булеву функцию  $F(x_1, \dots, x_n)$ . Требуется построить сокращенную дизъюнктивную нормальную форму (ДНФ) функции  $F$ , т.е. ДНФ, состоящую из всех максимальных конъюнкций функции  $F$ .

Покажем, что монотонная дуализация — это задача построения минимальных независимых от  $R$  элементов множества  $P$  при условии, что  $P$  —  $n$ -мерный булев куб,  $P_i = \{0, 1\}$  при  $i = 1, 2, \dots, n$ , в каждом  $P_i$  задан порядок  $0 < 1$  и  $R$  — множество нулей функции  $F$ , содержащее множество так называемых “верхних нулей” функции  $F$ .

Действительно, пусть  $E_0$  и  $E_1$  — множества вершин  $n$ -мерного булева куба  $P$ , на которых функция  $F$  принимает соответственно значение 0 и 1. Множества  $I(E_1)$  и  $I(E_0)$  называются соответственно верхними нулями и нижними единицами функции  $F$ . Каждое из этих множеств является антицепью. Очевидно, что  $E_0 = (I(E_1))^-$ ,  $E_1 = (I(E_0))^+$ . Представление монотонной булевой функции в виде КНФ равносильно заданию множества нулей этой функции, содержащего множество ее верхних нулей. Задание множества нижних единиц монотонной булевой функции равносильно заданию ее в виде сокращенной ДНФ. Таким образом, монотонная дуализация — это задача построения  $I(R^-)$  при  $I(E_1) \subseteq R \subseteq E_0$  (в частности, это задача построения нижних единиц монотонной булевой функции, заданной верхними нулями).

Заметим, что если в искомой сокращенной ДНФ функции  $F$  заменить знаки  $\&$  и  $\vee$  соответственно на знаки  $\vee$  и  $\&$ , то получим КНФ функции  $dual F$ . По определению  $dual F = \bar{F}(\bar{\alpha}_1, \dots, \bar{\alpha}_n)$  для любого набора  $(\alpha_1, \dots, \alpha_n)$ . Отсюда название задачи.

Задача монотонной дуализации может быть решена, во-первых, на основе перечисления неприводимых покрытий булевой матрицы, строками которой являются наборы из  $R$ , и, во-вторых, на основе перечисления минимальных вершинных покрытий гиперграфа с  $n$  вершинами и  $|R|$  ребрами [6], [12], [14], [17].

Теоретические оценки эффективности алгоритмов дуализации базируются на оценке сложности одного шага, т.е. сложности поиска одного нового решения. Наиболее эффективным считается алгоритм, имеющий полиномиальный от размера входа шаг. Такой алгоритм называется алгоритмом с полиномиальной задержкой [14]. Однако полиномиальные алгоритмы удалось построить лишь для некоторых частных случаев монотонной дуализации, например, для случая 2-КНФ. В настоящее время сформировались два основных направления исследований.

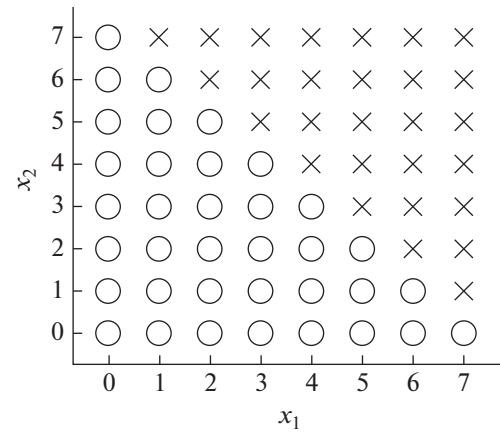
Первое направление нацелено на построение так называемых инкрементальных алгоритмов, когда алгоритму разрешено просматривать решения, найденные на предыдущих шагах. При этом оценка сложности шага алгоритма дается для худшего случая (для самого сложного варианта задачи). В [18] построен инкрементальный алгоритм дуализации гиперграфа с квазиполиномиальным шагом, определяемым фактически не только размером входа задачи, но и размером ее выхода. В [19] для случая, когда каждое  $P_i$ ,  $i = 1, 2, \dots, n$ , является цепью и  $|P_i| \geq 2$ , на базе алгоритма, предложенного в [18], построен квазиполиномиальный инкрементальный алгоритм. Подход интересен в основном для теории, поскольку в худшем случае число решений дуализации (размер выхода задачи) растет экспоненциально с ростом размера ее входа.

Второе направление основано на построении асимптотически оптимальных алгоритмов дуализации. В этом случае алгоритму разрешено делать лишние полиномиальные шаги при условии, что их число почти всегда должно быть достаточно мало по сравнению с числом всех решений задачи. В рамках данного направления удалось построить алгоритмы монотонной дуализации, эффективные в типичном случае (эффективные для почти всех вариантов задачи). Эти алгоритмы имеют теоретическое обоснование и являются лидерами по скорости счета [11]–[13].

В [20]–[22] предложен асимптотически оптимальный алгоритм построения  $I(R^+)$  для случая, когда каждое множество  $P_i = \{0, 1, \dots, k-1\}$ ,  $i = 1, 2, \dots, n$ , — цепь и элементы в  $P_i$  упорядочены по



Фиг. 1.



Фиг. 2.

возрастанию. При обосновании асимптотической оптимальности данного алгоритма используется матричная формулировка дуализации над произведением цепей.

### 3. ЛОГИЧЕСКАЯ КЛАССИФИКАЦИЯ ЧАСТИЧНО УПОРЯДОЧЕННЫХ ДАННЫХ

#### 3.1. Постановка задачи и основные утверждения

Как уже было отмечено во введении, решение прикладных задач классификации не всегда возможно в рамках классической постановки логической классификации, не учитывающей наличие сложных отношений на множествах допустимых значений признаков. В данном разделе предложена более общая постановка логической классификации, нацеленная на решение задач, в которых каждый признак принимает значения из некоторого конечного частично упорядоченного множества чисел. В этой новой постановке дано описание рассмотренных выше классических моделей.

На фиг. 1 и 2 представлены модельные примеры, наглядно демонстрирующие преимущество более общей постановки логической классификации. Рассматриваются две задачи классификации, каждая из которых с двумя классами и системой признаков  $\{x_1, x_2\}$ . Обучающие объекты классов  $K_1$  и  $K_2$  изображены соответственно в виде крестиков и в виде кружочков. Множество допустимых значений каждого признака – множество целых чисел.

Рассмотрим первый пример (фиг. 1). Нетрудно видеть, что для любого значения признака  $x_1$  число объектов каждого класса, у которых признак  $x_1$  принимает данное значение, равно четырем. То же верно и для признака  $x_2$ . Также нетрудно видеть, что каждый эл.кл. ранга 2 является корректным либо для класса  $K_1$ , либо для класса  $K_2$  и что только один объект из обучающей выборки его порождает. Следовательно, любой корректный эл.кл. любого из двух классов обладает низкой информативностью. Такой эл.кл. не порождается ни одним объектом, не входящим в обучающую выборку. Схожая ситуация и со вторым примером (фиг. 2). Существует всего два корректных эл.кл. ранга 1, а именно, эл.кл.  $((0), \{x_1\})$  и  $((0), \{x_2\})$ , и каждый из них порождается лишь несколькими объектами класса  $K_2$ .

С другой стороны, если на множестве допустимых значений признаков  $x_1$  и  $x_2$  ввести естественный порядок  $\dots \preceq 0 \preceq 1 \preceq \dots$ , то решающие правила для обоих примеров становятся простыми. Пусть  $S = (a_1, a_2)$  – распознаваемый объект. Тогда решающее правило для первого примера: если  $(a_1, a_2) \preceq (3, 3)$  или  $(a_1, a_2) \succeq (4, 4)$ , то  $S \in K_1$ , иначе  $S \in K_2$ . Решающее правило для второго примера: если  $(a_1, a_2) \preceq (x, 7 - x)$  для хотя бы одного  $x \in \{0, 1, \dots, 7\}$ , то  $S \in K_2$ , иначе  $S \in K_1$ .

Рассмотрим задачу классификации по прецедентам при условии, что на множестве допустимых значений каждого признака задан частичный порядок.

Пусть  $M = N_1 \times \dots \times N_n$ , где  $N_i, i \in \{1, 2, \dots, n\}$ , – конечное множество значений признака  $x_i$ , на котором задан частичный порядок. Будем считать, что каждое множество  $N_i, i \in \{1, 2, \dots, n\}$ , имеет



наименьший и наибольший элементы, т.е. такие элементы  $q_i$  и  $k_i$ , для которых выполнено  $a \succcurlyeq q_i$  и  $a \preccurlyeq k_i$  для любого  $a \in N_i$ . Если наименьший (наибольший) элемент в  $N_i$  отсутствует, то  $N_i$  дополним таким элементом.

Близость объекта  $S = (a_1, \dots, a_n)$  из  $M$  и эл.кл.  $(\sigma, H)$ ,  $H = \{x_{j_1}, \dots, x_{j_r}\}$ ,  $\sigma_i = (\sigma_{i_1}, \dots, \sigma_{i_r})$ ,  $\sigma_i \in N_{j_i}$ , при  $i = 1, 2, \dots, r$ , будем оценивать величиной  $\hat{B}(\sigma, S, H)$ , равной 1, если  $a_{j_i} \preccurlyeq \sigma_i$  при  $i = 1, 2, \dots, r$ , и равной 0 в противном случае. Будем говорить, что объект  $S$  порождает эл.кл.  $(\sigma, H)$ , если  $\hat{B}(\sigma, S, H) = 1$ .

Данные в разд. 1 понятия корректного эл.кл. класса  $K$ , представительного эл.кл. класса  $K$ , покрытия класса  $K$  и теста полностью переносятся на рассматриваемый случай, если  $B(\sigma, S, H)$  заменить на  $\hat{B}(\sigma, S, H)$ .

Пусть  $(\sigma, H)$  – эл.кл., в котором  $H = \{x_{j_1}, \dots, x_{j_r}\}$ ,  $\sigma = (\sigma_1, \dots, \sigma_r)$ ,  $\sigma_i \in N_{j_i}$ ,  $i = 1, 2, \dots, r$ . Эл.кл.  $(\sigma, H)$  сопоставим набор  $S_{(\sigma, H)} = (\gamma_1, \dots, \gamma_n)$  из  $M = N_1 \times \dots \times N_n$ , в котором  $\gamma_t = \sigma_i$  при  $t = j_i$ ,  $i = 1, 2, \dots, r$ , и  $\gamma_t = k_t$  при  $t \notin \{j_1, \dots, j_r\}$ .

Покрытие  $(\sigma, H)$  класса  $K$  назовем *тупиковым*, если любой эл.кл.  $(\sigma', H')$  такой, что  $S_{(\sigma, H)} \prec S_{(\sigma', H')}$ , не является покрытием класса  $K$ .

Через  $R(K)$  обозначим множество прецедентов из класса  $K$ . Нетрудно убедиться в справедливости приводимых ниже утверждений 2 и 3.

**Утверждение 2.** Эл.кл.  $(\sigma, H)$  является покрытием класса  $K$  относительно заданного частичного порядка тогда и только тогда, когда  $S_{(\sigma, H)} \in M/R(K)^+$ .

**Утверждение 3.** Покрытие  $(\sigma, H)$  класса  $K$  является тупиковым покрытием класса  $K$  относительно заданного частичного порядка тогда и только тогда, когда  $S_{(\sigma, H)} \in I(R(K)^+)$ .

Представительный для класса  $K$  эл.кл.  $(\sigma, H)$  назовем *тупиковым*, если любой эл.кл.  $(\sigma', H')$  такой, что  $S_{(\sigma, H)} \prec S_{(\sigma', H')}$ , не является представительным для класса  $K$ .

Пусть  $\bar{K} = M \setminus K$ . Будем рассматривать  $\bar{K}$  как отдельный класс, т.е. будем считать, что есть всего два класса  $K$  и  $\bar{K}$ . Тогда очевидным является

**Утверждение 4.** Эл.кл.  $(\sigma, H)$  является тупиковым представительным для класса  $K$  относительно заданного частичного порядка тогда и только тогда, когда  $S_{(\sigma, H)} \in I(R(\bar{K})^+)$  и  $S_{(\sigma, H)} \in R(K)^+$ .

Из утверждений 3 и 4 следует, что в случае частично упорядоченных данных при построении логических классификаторов, основанных на голосовании по тупиковым покрытиям класса и тупиковым представительным эл.кл. класса, возникает задача построения максимальных независимых элементов произведения частичных порядков.

Заметим, что если описание обучающего объекта  $S$  класса  $K$  следует за описанием некоторого объекта из  $\bar{K}$ , то, очевидно,  $S$  не порождает ни одного представительного эл.кл. класса  $K$ . Поэтому в общем случае существование представительных для класса  $K$  эл.кл. не гарантировано и для корректности алгоритма голосования по представительным эл.кл. или тестам необходимо, чтобы описания объектов из разных классов были несравнимы.

Покажем, что в случае непересекающихся классов существует такое преобразование признакового описания множества  $M$ , в результате которого для каждого класса  $K$  всегда образуется непустое множество тупиковых представительных эл.кл. и любой прецедент из  $K$  порождает хотя бы один эл.кл. из данного множества.

Обозначим через  $\tilde{P}$  множество, совпадающее с множеством  $P$ , но с обратным отношением порядка, т.е.  $x \preccurlyeq y$  в  $P$  тогда и только тогда, когда  $y \preccurlyeq x$  в  $\tilde{P}$ .

Пусть  $\tilde{M} = \tilde{N}_1 \times \dots \times \tilde{N}_n$ . Зададим отображение  $\varphi: M \rightarrow M \times \tilde{M}$  следующим образом. Отображение  $\varphi$  переводит объект  $S = (a_1, \dots, a_n)$  из  $M$  в объект  $\varphi(S) = (a_1, \dots, a_n, a_{n+1}, \dots, a_{2n})$  из  $M \times \tilde{M}$ , в котором  $a_{i+n} = a_i$  при  $i \in \{1, 2, \dots, n\}$ . Иными словами, признаковое описание объекта  $S$  дублируется с обратным отношением порядка.

Пусть  $\varphi(A)$ ,  $A \subset M$ , – образ  $A$  при отображении  $\varphi$ . Имеет место следующая

**Теорема 1.** Если классы множества  $M$  не пересекаются, то любой прецедент из класса  $\varphi(K)$  порождает тупиковый представительный эл.кл. класса  $\varphi(K)$ .

**Доказательство.** Пусть  $S \in R(K)$ . Покажем, что элемент  $\varphi(S)$  не принадлежит  $\varphi(R(\bar{K}))^+$  (независим от  $\varphi(R(\bar{K}))$  в  $\varphi(M)$ ). Действительно, пусть  $\varphi(\bar{S}) \preceq \varphi(S)$ ,  $\bar{S} \in R(\bar{K})$ . Тогда имеем:  $\bar{S} \preceq S$  и  $S \preceq \bar{S}$ . Откуда в силу антисимметричности заданного отношения частичного порядка получаем  $\bar{S} = S$ , что противоречит условию  $R(K) \cap R(\bar{K}) = \emptyset$ . В силу конечности  $M \times \tilde{M}$  существует максимальный независимый от  $\varphi(R(\bar{K}))$  элемент  $\varphi(S') \in M \times \tilde{M}$  такой, что  $\varphi(S) \preceq \varphi(S')$ . Отсюда следует утверждение теоремы.

**Замечание 1.** Отметим, что если каждое множество  $N_i$ ,  $i \in \{1, 2, \dots, n\}$ , – антицепь с добавленным наибольшим элементом, то  $R(K)^+ = R(K)$  и  $\hat{B}(\sigma, S, H) = B(\sigma, S, H)$  и, следовательно, в этом случае мы имеем классическую постановку, рассмотренную в разд. 1.

### 3.2. О построении $I(R^+)$

В [20]–[22] дана матричная постановка задачи дуализации над произведением конечных цепей и построен асимптотически оптимальный алгоритм, решающий данную задачу путем перечисления так называемых упорядоченных тупиковых покрытий целочисленной матрицы.

Ниже приведена матричная формулировка задачи построения максимальных независимых элементов для произведения конечных частичных порядков  $P$ , опирающаяся на понятие упорядоченного тупикового покрытия матрицы, строками которой являются наборы из  $R \subseteq P$ .

Введем обозначения:  $Q_1(x, P)$ ,  $x \in P$ , – множество всех элементов в  $P$ , непосредственно следующих за  $x$  ( $Q_1(x, P) = \{y \in P: x < y, \forall a \in P: x < a \Rightarrow a \not\prec y\}$ );  $Q_2(x, y, P)$ ,  $x \in P$ ,  $y \in Q_1(x, P)$ , – множество всех элементов  $a \in P$ , не предшествующих  $x$  и предшествующих  $y$  ( $Q_2(x, y, P) = \{a \in P: a \not\prec x, a \preceq y\}$ ).

Элемент  $x$  частично упорядоченного множества  $P$  называется *максимальным элементом* в  $P$ , если  $Q_1(x, P) = \emptyset$ , т.е. для любого  $y \in P$  либо  $y \preceq x$ , либо  $x$  и  $y$  несравнимы.

Пусть  $P = P_1 \times \dots \times P_n$  и каждое  $P_i$ ,  $i \in \{1, 2, \dots, n\}$ , – конечное частично упорядоченное множество с наибольшим элементом. В заданных обозначениях введем понятие упорядоченного тупикового покрытия матрицы  $L_R$ , строками которой являются наборы из  $R \subseteq P$ .

Пусть  $H$  – набор столбцов матрицы  $L_R$  с номерами  $j_1, \dots, j_r$ ,  $\sigma = (\sigma_1, \dots, \sigma_r)$ ,  $\sigma_i \in P_{j_i}$  при  $i = 1, 2, \dots, r$ . Тогда набор столбцов  $H$  называется *упорядоченным тупиковым  $\sigma$ -покрытием*, если выполнены следующие два условия: 1) для любого  $i \in \{1, 2, \dots, r\}$  и любого  $y \in Q_1(\sigma_i, P_{j_i})$  подматрица  $L_R^H$  матрицы  $L_R$ , образованная столбцами из  $H$ , содержит каждую из строк вида  $(\beta_1, \dots, \beta_{i-1}, \beta_i, \beta_{i+1}, \dots, \beta_r)$ , где  $\beta_r \in Q_2(\sigma_i, y, P_{j_i})$  и  $\beta_t \preceq \sigma_t$  при  $t \neq i$ ,  $t \in \{1, 2, \dots, r\}$ ; 2) подматрица  $L_R^H$  не содержит строк, предшествующих  $\sigma$ .

Заметим, что если  $P_i$ ,  $i \in \{1, 2, \dots, n\}$ , – конечная цепь и  $x \in P_i$  не является наибольшим элементом в  $P_i$ , то множество  $Q_1(x, P_i)$  состоит из одного элемента, обозначаемого далее через  $x + 1$ , и следовательно,  $Q_2(x, x + 1, P_i) = \{x + 1\}$ . Поэтому в случае произведения конечных цепей условие 1) из определения упорядоченного тупикового  $\sigma$ -покрытия превращается в следующее условие: для любого  $i \in \{1, 2, \dots, r\}$  подматрица  $L_R^H$  матрицы  $L_R$ , образованная столбцами из  $H$ , содержит строку  $(\beta_1, \dots, \beta_{i-1}, \sigma_i + 1, \beta_{i+1}, \dots, \beta_r)$ , где  $\beta_t \preceq \sigma_t$  при  $t \neq i$ ,  $t \in \{1, 2, \dots, r\}$ .

Покажем, что между множеством упорядоченных тупиковых покрытий матрицы  $L_R$  и множеством  $I(R^+)$  существует взаимно однозначное соответствие.

Возьмем элемент  $x = (x_1, \dots, x_n) \in P$ , в котором компонента  $x_{j_i} = \sigma_i$ ,  $i \in \{1, 2, \dots, r\}$ , не является наибольшим элементом в  $P_{j_i}$ , а каждая из остальных компонент  $x_j$ ,  $j \in \{1, 2, \dots, n\} \setminus \{j_1, \dots, j_r\}$  – наибольший элемент в  $P_j$ . Положим  $\sigma = (\sigma_1, \dots, \sigma_r)$ . Имеет место следующая

**Теорема 2.** *Элемент  $x$  является максимальным независимым от  $R$  тогда и только тогда, когда набор столбцов матрицы  $L_R$  с номерами  $j_1, \dots, j_r$  является упорядоченным тупиковым  $\sigma$ -покрытием матрицы  $L_R$ .*

**Доказательство.** Достаточность очевидна. Докажем необходимость.

Пусть элемент  $x$  является максимальным независимым от  $R$ . Тогда из условия независимости элемента  $x$  следует условие 2) из определения упорядоченного тупикового  $\sigma$ -покрытия матрицы  $L_R$ . Покажем справедливость условия 1) из того же определения.

Так как при  $i \in \{1, 2, \dots, r\}$  компонента  $x_{j_i}$  элемента  $x$ , не является наибольшим элементом в  $P_{j_i}$ , то  $Q_1(\sigma_i, P_{j_i}) \neq \emptyset$  для всех  $i \in \{1, 2, \dots, r\}$ . Пусть  $i \in \{1, 2, \dots, r\}$ ,  $\sigma'_i \in Q_1(\sigma_i, P_{j_i})$ . Рассмотрим элемент  $x' = (x'_1, \dots, x'_n)$  множества  $P$ , полученный из элемента  $x$  путем замены компоненты  $x_{j_i}$ , равной  $\sigma_i$ , на компоненту  $x'_{j_i}$ , равную  $\sigma'_i$ . Из того, что  $x$  – максимальный независимый от  $R$  элемент, следует, что  $x' \in R^+$ . Это означает, что в  $R$  существует элемент  $x'' = (x''_1, \dots, x''_n)$ , такой что  $x'' \preceq x'$ . Положим  $x''_{j_t} = \beta_t$ ,  $t \in \{1, 2, \dots, r\}$ . Тогда из сказанного следует, что  $\beta_t \preceq \sigma'_t$ ,  $\sigma'_t = \sigma_t$  при  $t \neq i$ ,  $t \in \{1, 2, \dots, r\}$ , и  $\beta_i \preceq \sigma'_i$ ,  $\sigma'_i \in Q_1(\sigma_i, P_{j_i})$ . Если  $\beta_i \preceq \sigma_i$ , то  $x'' \preceq x$ , что противоречит независимости  $x$ . Значит,  $\beta_i \not\preceq \sigma_i$ , т.е.  $\beta_i \in Q_2(\sigma_i, \sigma'_i, P_{j_i})$ . Таким образом, в силу произвольности выбора компоненты  $\sigma_i$  элемента  $x$ , а также произвольности выбора элемента  $\sigma'_i \in Q_1(\sigma_i, P_{j_i})$ , условие 1) оказывается выполненным и действительно набор столбцов матрицы  $L_R$  с номерами  $j_1, \dots, j_r$  является упорядоченным тупиковым  $\sigma$ -покрытием матрицы  $L_R$ .

Теорема доказана.

**Замечание 2.** Согласно теореме 2 логическая классификация данных, являющихся произведением конечных частичных порядков, фактически сводится к поиску упорядоченных тупиковых покрытий целочисленной матрицы.

**Замечание 3.** В разд. 2 было отмечено, что задача монотонной дуализации может быть решена на основе перечисления неприводимых покрытий булевой матрицы. Введенное в данном разделе понятие упорядоченного тупикового покрытия матрицы  $L_R$  обобщает понятие неприводимого покрытия булевой матрицы. Действительно, если  $R \subseteq P$ ,  $P = P_1 \times \dots \times P_n$ ,  $P_i = \{0, 1\}$ ,  $i \in \{1, 2, \dots, n\}$ , и в каждом  $P_i$  задан порядок  $0 < 1$ , то упорядоченное тупиковое  $(0, \dots, 0)$ -покрытие матрицы  $L_R$  является неприводимым покрытием.

## СПИСОК ЛИТЕРАТУРЫ

1. *Бонгард М.М.* Проблема узнавания. М.: Физматгиз, 1967. 321 с.
2. *Вайнцвайг М.Н.* Алгоритм обучения распознаванию образов “Кора” // Алгоритмы обучения распознаванию образов / Под ред. В.Н. Вапник. М.: Сов. радио, 1973. С. 110–116.
3. *Дмитриев А.И., Журавлёв Ю.И., Кренделев Ф.П.* О математических принципах классификации предметов или явлений // Дискретный анализ. Новосибирск: ИМ СО АН СССР, 1966. Вып. 7. С. 3–17.
4. *Баскакова Л.В., Журавлёв Ю.И.* Модель распознающих алгоритмов с представительными наборами и системами опорных множеств // Ж. вычисл. матем. и матем. физ. 1981. Т. 21. № 5. С. 1264–1275.
5. *Дюкова Е.В., Песков Н.В.* Поиск информативных фрагментов описаний объектов в дискретных процедурах распознавания // Ж. вычисл. матем. и матем. физ. 2002. Т. 42. № 5. С. 741–753.
6. *Дюкова Е.В., Песков Н.В.* Построение распознающих процедур на базе элементарных классификаторов // Математические вопросы кибернетики. 2005. № 14. С. 57–92. URL: [www.ccas.ru/frc/papers/djukova05construction.pdf](http://www.ccas.ru/frc/papers/djukova05construction.pdf)
7. *Дюкова Е.В., Журавлёв Ю.И., Рудаков К.В.* Об алгебраическом синтезе корректирующих процедур распознавания на базе элементарных алгоритмов // Ж. вычисл. матем. и матем. физ. 1996. Т. 36. № 8. С. 217–225.
8. *Дюкова Е.В., Журавлёв Ю.И., Сотнезов Р.М.* Построение коллектива логических корректоров на базе элементарных классификаторов // J. Pattern Recognition and Image Analysis. 2011. Vol. 21. № 2.
9. *Дюкова Е.В., Журавлёв Ю.И., Прокофьев П.А.* Логические корректоры в задаче классификации по прецедентам // Ж. вычисл. матем. и матем. физ. 2017. Т. 57. № 11. С. 1906–1927.
10. *Murakami K., Uno T.* Efficient algorithms for dualizing large-scale hypergraphs // Discrete Applied Mathematics. 2014. V. 170. P. 83–94.

11. Дюкова Е.В. Об асимптотически оптимальном алгоритме построения тупиковых тестов // Докл. АН СССР. 1977. Т. 233. № 4. С. 527–530.
12. Дюкова Е.В. О сложности реализации некоторых процедур распознавания // Ж. вычисл. матем. и матем. физ. 1987. Т. 27. № 1. С. 114–127.
13. Дюкова Е.В., Прокофьев П.А. Об асимптотически оптимальных алгоритмах дуализации // Ж. вычисл. матем. и матем. физ. 2015. Т. 55. № 5. С. 895–910.
14. Johnson D.S., Yannakakis M., Papadimitriou C.H. On general all maximal independent sets // Information Processing Letters. 1988. V. 27. Iss. 3. P. 119–123.
15. Djukova E.V., Peskov N.V. A Classification algorithm based on the complete decision tree // J. Pattern Recognition and Image Analysis. 2007. V. 17. N. 3. P. 363–367.
16. Журавлёв Ю.И. Об алгебраическом подходе к решению задач распознавания или классификации // Пробл. кибернетики. М.: Наука, 1978. Вып. 33. С. 5–68.
17. Дюкова Е.В., Журавлёв Ю.И. Дискретный анализ признаковых описаний в задачах распознавания большой размерности // Ж. вычисл. матем. и матем. физ. 2000. Т. 40. № 8. С. 1264–1278.
18. Fredman L., Khachiyan L.. On the complexity of dualization of monotone disjunctive normal forms. Journal of Algorithms, 21:618–628, 1996.
19. Boros E., Elbassioni K., Gurvich V., Khachiyan L., Makino K. Dual-bounded generating problems: All minimal integer solutions for a monotone system of linear inequalities. SIAM Journal on Computing, 31(5): 1624–1643, 2002.
20. Дюкова Е.В., Масляков Г.О., Прокофьев П.А. О дуализации над произведением частичных порядков // Машинное обучение и анализ данных. 2017. Т. 3. № 4. С. 239–249. <http://jmla.org/papers/doc/2017/no4/Djukova2017Dualization.pdf>
21. Дюкова Е.В., Масляков Г.О., Прокофьев П.А. Дуализация над произведением цепей: асимптотические оценки числа решений // Докл. АН. 2018. Т. 483. № 2. С. 130–133.
22. Dyukova E.V., Maslyakov G.O., Prokof'ev P.A. Computational Mathematics and Modeling. 2019. Vol. 30. No. 1. January, 2019. P. 7–12.