

## КЛАССИФИКАЦИЯ ПО ПРЕЦЕДЕНТАМ: НОВЫЕ МОДЕЛИ ЛОГИЧЕСКИХ КОРРЕКТОРОВ<sup>1,2</sup>

© 2025 г. И. Е. Генрихов<sup>a</sup>, Е. В. Дюкова<sup>a, \*</sup>

<sup>a</sup>Федеральный исследовательский центр “Информатика и управление”  
Российской академии наук, Москва, Россия

\*e-mail: edjukova@mail.ru

Поступила в редакцию 21.11.2024 г.

После доработки 28.11.2024 г.

Принята к публикации 13.01.2025 г.

Рассматривается подход к задаче корректной классификации по прецедентам, основанный на применении логических методов анализа данных. Приводится описание схемы работы моделей логических классификаторов, нацеленных на построение специальных фрагментов описаний прецедентов, называемых корректными элементарными классификаторами. Более сложные модели, а именно модели логических корректоров, базируются на синтезе семейств корректных наборов элементарных классификаторов. В отличие от классических моделей логические корректоры показывают хорошие результаты в случае многозначных признаков, т.е. признаков, имеющих большое число различных значений. Изучаются вопросы сокращения временных затрат и повышения качества классификации логических корректоров. Предлагаются новые детерминированные и стохастические варианты таких моделей, предназначенные для работы с частично упорядоченными данными. Приводятся результаты экспериментов на модельных и реальных данных.

**Ключевые слова:** корректная классификация по прецедентам, элементарный классификатор, корректный набор элементарных классификаторов, монотонный логический корректор, неприводимое покрытие булевой матрицы, корректирующее покрытие целочисленной матрицы

DOI: 10.31857/S0002338825010053 EDN: AGYDCM

## SUPERVISED CLASSIFICATION PROBLEM: NEW MODELS OF LOGICAL CORRECTORS

I. E. Genrikhov<sup>a</sup>, E. V. Djukova<sup>a, \*</sup>

<sup>a</sup>Federal Research Center “Computer Science and Control”  
of the Russian Academy of Sciences, Moscow, Russia

\*e-mail: edjukova@mail.ru

An approach to the correct supervised classification problem based on the application of logical data analysis methods is considered. The description of the operation scheme of logical classifier models aimed at constructing special fragments of precedent descriptions, called correct elementary classifiers, is provided. More complex models, namely models of logical correctors, are based on the synthesis of families of correct sets of elementary classifiers. Unlike classical models, logical correctors show good results in the case of multivalued features, i.e. features with a large number of different values. The article examines issues related to reducing time costs and improving the quality of classification of logical correctors. New deterministic and stochastic variants of such models are proposed, designed to work with partially ordered data. The results of experiments on model and real data are presented.

**Keywords:** correct supervised classification, elementary classifier, correct set of elementary classifiers, monotone logical corrector, irreducible covering of a Boolean matrix, corrective covering of an integer matrix

<sup>1</sup> Исследование выполнено за счет гранта Российского научного фонда № 24-21-00301, <https://rscf.ru/project/24-21-00301/>.

<sup>2</sup> Статья рекомендована к публикации программным комитетом конференции ИОИ-2024 (23–27 сентября 2024 г., г. Гродно) по результатам доклада.

**Введение.** Задача корректной классификации на основе прецедентов рассматривается в следующей постановке.

Исследуется множество объектов  $M$ . Каждый объект этого множества может быть представлен в виде числового набора, полученного на основе наблюдения или измерения ряда его характеристик  $x_1, \dots, x_n$ , называемых признаками. Предполагается, что каждый признак имеет ограниченное множество допустимых значений, которые кодируются целыми числами. Известно, что  $M$  представимо в виде объединения непересекающихся подмножеств  $K_1, \dots, K_l$ ,  $l \geq 2$ , называемых классами. Имеется конечный набор объектов множества  $M$ , про каждый из которых известно, какому классу он принадлежит. Это прецеденты, или обучающие объекты. Требуется по предъявленному набору значений признаков, описывающему некоторый объект из  $M$ , определить (распознать) класс, которому принадлежит этот объект.

Логический или дискретный подход к задаче классификации по прецедентам возник в связи с необходимостью прогнозировать редкие события, для которых нет достаточного статистического материала. Фундаментальную роль в создании отечественных методов логической классификации сыграли работы член-корр. РАН С.В. Яблонского, в которых введено хорошо известное в дискретной математике понятие теста [1], и работы акад. РАН Ю.И. Журавлева, опубликованные в 70–80-х гг. прошлого века. Дальнейшее развитие направление получило в большом числе публикаций, среди которых статьи [2–4] имеют непосредственное отношение к исследованиям, проводимым в данной статье.

При конструировании логических классификаторов большое внимание уделяется вопросам синтеза корректных алгоритмов, т.е. алгоритмов, не ошибающихся на обучающей выборке. Предполагается, что каждый признак имеет ограниченное множество допустимых значений, которые кодируются целыми числами. Обучение классификатора сводится к поиску в исходных данных информативных фрагментов описаний прецедентов. Такие фрагменты позволяют различать объекты из разных классов и, 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. В случае большого числа признаков возникают вычислительные трудности, связанные с необходимостью решать сложные дискретные задачи.

Основным понятием логического подхода является понятие элементарного классификатора. Пусть  $N_j, j=1, n$ , – множество допустимых значений признака  $x_j$ ;  $H$  – набор из  $r$  различных признаков вида  $H = \{x_{j_1}, \dots, x_{j_r}\}$ ;  $\sigma = (\sigma_1, \dots, \sigma_r)$  – набор, в котором  $\sigma_i \in N_{j_i}$  при  $i=1, r$ . Пара  $(\sigma, H)$  называется элементарным классификатором (ЭК) ранга  $r$ . ЭК  $(\sigma, H)$  может быть задан в виде элементарной конъюнкции  $B_{(\sigma, H)} = x_{j_1}^{\sigma_1} \& \dots \& x_{j_r}^{\sigma_r}$  над переменными  $x_1, \dots, x_n$ . Если объект  $S$  из  $M$  имеет признаконое описание  $(a_1, \dots, a_n)$ , в котором  $a_j, j=1, n$ , – значение признака  $x_j$  для объекта  $S$ , и  $a_{j_i} = \sigma_i$  при  $i=1, r$ , то говорят, что ЭК  $(\sigma, H)$  содержится в объекте  $S$ . В этом случае интервал истинности конъюнкции  $B_{(\sigma, H)}$  содержит  $S$ . ЭК называется корректным для класса  $K$ ,  $K \in \{K_1, \dots, K_l\}$ , если не существует двух прецедентов, содержащих данный ЭК, при этом один из этих прецедентов принадлежит  $K$ , а второй не принадлежит  $K$ .

В классических моделях в процессе обучения строятся семейства корректных ЭК. На этапе классификации каждый найденный ЭК участвует в процедуре голосования и формирования оценок принадлежности распознаваемого объекта к классам. Трудными являются задачи, в которых практически каждый корректный ЭК содержится в небольшом числе прецедентов. Подобная ситуация возникает, например, в случае, когда признаки имеют много значений. В работе [2] предложены модели логических корректоров, в которых используются идеи алгебраического подхода [5–7] к синтезу корректных классификаторов. В данном случае в качестве базовых алгоритмов рассматриваются произвольные ЭК, а в качестве корректирующих функций – булевы функции.

Логический корректор основан на построении семейства корректных наборов ЭК. Корректный для класса  $K$  набор ЭК позволяет различать любую пару прецедентов, один из которых принадлежит классу  $K$ , а другой не принадлежит этому классу, поскольку этот набор содержит ЭК  $(\sigma, H)$ , обладающий свойством: конъюнкция  $B(\sigma, H)$  принимает разные значения на данных объектах. Более эффективны модели логических корректоров, использующие в качестве корректирующей функции монотонную булеву функцию. Каждый монотонный корректор класса  $K$  рассматривает только те ЭК, которые содержатся в прецедентах из  $K$ .

Основная проблема, возникающая при синтезе монотонного логического корректора класса  $K$ , – это вычислительная сложность этапа обучения, на котором решается задача поиска минимальных покрытий булевой матрицы  $L_K$  значительных размеров. Каждая строка этой матрицы порождается парой прецедентов, один из которых принадлежит  $K$ , а другой не принадлежит  $K$ . Число столбцов в  $L_K$  равно числу тех ЭК, каждый из которых содержитя в хотя бы одном обучающем объекте класса  $K$ . В простейшем случае, а именно когда признаки бинарные и рассматриваются только ЭК ранга 1, число столбцов в  $L_K$  равно  $2n$ , где  $n$  – число признаков. В работе [3] построена и исследована модель логического корректора с монотонными корректными наборами ЭК ранга 1 и с применением генетического алгоритма для поиска минимальных покрытий матрицы  $L_K$ .

Для сокращения временных затрат в случае бинарных данных в работе [2] предложено осуществлять поиск монотонных корректных наборов из ЭК ранга 1 на основе построения специального вида покрытий булевой матрицы с тем же числом строк, что и в  $L_K$ , и с числом столбцов, равным числу признаков  $n$ . Однако в работе [2] не было приведено экспериментальных данных, подтверждающих существенное сокращение перебора. Эта идея обобщена в настоящей работе на случай целочисленных данных. Проведенные исследования показали эффективность подхода. Рассмотрены вопросы разработки стохастических моделей логических корректоров и модификации этих моделей на случай, когда на множествах значений признаков заданы конечные частичные порядки, т.е. признаковые описания прецедентов – это элементы декартона произведения конечных частичных порядков. Классические модели логических классификаторов над произведением частичных порядков впервые построены в работе [8].

**1. Логические классификаторы: принципы работы и основные модели.** Среди классических моделей логических классификаторов хорошие результаты на данных с малой значностью признаков показывают модели, в которых в качестве корректных ЭК рассматриваются представительные ЭК классов.

ЭК  $(\sigma, H)$  называется представительным для класса  $K$ ,  $K \in \{K_1, \dots, K_l\}$ , если  $(\sigma, H)$  содержитя в хотя бы одном прецеденте из  $K$  и не содержитя ни в каком прецеденте не из  $K$ . Понятие представительного ЭК введено в работе [9]. Согласно общей схеме работы рассматриваемых моделей [10], на этапе обучения задается некоторый частичный порядок на множестве  $\mathcal{P}(K)$  представительных ЭК класса  $K$  и ищутся максимальные относительно заданного порядка элементы этого множества.

Пусть на  $\mathcal{P}(K)$  установлен порядок, согласно которому ЭК  $(\sigma', H')$  следует за ЭК  $(\sigma, H)$ , если интервал истинности конъюнкции  $B_{(\sigma', H')}$  содержит интервал истинности конъюнкции  $B_{(\sigma, H)}$ . Иначе  $(\sigma', H')$  и  $(\sigma, H)$  несравнимы. Тогда максимальные элементы частично упорядоченного множества  $\mathcal{P}(K)$  называются тупиковыми представительными ЭК.

Решающее правило основано на процедуре “голосования”, в которой участвуют найденные представительные ЭК классов [11]. Процедура “голосования” заключается в следующем. Если представительный ЭК  $(\sigma, H)$  класса  $K$  содержитя в распознаваемом объекте  $S$ , то  $(\sigma, H)$  дает положительную оценку  $\Gamma_K(S, \sigma, H)$  за принадлежность  $S$  к классу  $K$ , которая равна числу прецедентов, содержащих  $(\sigma, H)$ . В противном случае  $\Gamma_K(S, \sigma, H) = 0$ . Оценки, поданные представительными ЭК класса  $K$  за принадлежность  $S$  к этому классу, суммируются, и суммарная оценка делится на число прецедентов класса  $K$ . Объект  $S$  относится к классу, получившему наибольшую оценку. Если таких классов несколько, то происходит отказ от распознавания объекта  $S$ .

Корректность алгоритма обеспечивается за счет корректности каждого участвующего в голосовании представительного ЭК. Основная задача этапа обучения – поиск информативных представительных ЭК. Практика показывает, что наиболее информативными являются представительные наборы небольшого ранга, в частности тупиковые. Задачи, в которых значность признаков высока, сложны для классических логических алгоритмов распознавания, поскольку каждый корректный ЭК содержитя в небольшом числе прецедентов. Существуют различные способы решения этой проблемы.

Один из этих способов заключается в выполнении “корректной” перекодировки исходных признаков с целью понизить их значность [12]. В результате после перекодирования объекты

из разных классов остаются различимыми. Основная проблема – выбор функционала, характеризующего качество перекодировки. Построение наилучшей в смысле качества распознавания корректной перекодировки – труднорешаемая оптимизационная задача. Другой способ основан на идеях алгебрологического подхода, который был предложен в работе [2]. Этот подход базируется на использовании произвольных ЭК (не обязательно являющихся корректными) и объединяет идеи логического и алгебраического подходов.

Алгебраический подход, развиваемый школой Ю.И. Журавлева, применяется тогда, когда требуется скорректировать работу нескольких распознающих алгоритмов, каждый из которых безошибочно классифицирует лишь часть обучающих объектов [5–7]. 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 алгебрологическом подходе говорят, когда каждый базовый алгоритм однозначно определяется некоторым ЭК и корректирующие функции являются булевыми. Следует отметить, что в ряде работ термин “логический корректор” используется также в случае, когда работа семейства некорректных эвристических процедур классификации корректируется логической функцией (например, см. [7]).

Основным понятием алгебрологического подхода является понятие корректного набора ЭК. Рассмотрим набор ЭК  $U = \{(\sigma_1, H_1), \dots, (\sigma_r, H_r)\}$ . Обозначим через  $B_{(\sigma_i, H_i)}(S)$ ,  $i = 1, r$ ,  $S \in M$ ,  $S = (a_1, \dots, a_n)$ , значение, которое принимает конъюнкция  $B_{(\sigma_i, H_i)}$  на наборе  $(a_1, \dots, a_n)$ . Бинарный вектор  $U(S) = \{B_{(\sigma_1, H_1)}(S), \dots, B_{(\sigma_r, H_r)}(S)\}$  – отклик набора  $U$  на объекте  $S$ . Набор  $U$  называется корректным для класса  $K$ , если не существует двух прецедентов с одинаковыми откликами на этом наборе и таких, что один из них принадлежит  $K$ , а другой не принадлежит  $K$ . Корректный для класса  $K$  набор ЭК называется тупиковым, если никакое его собственное подмножество не является корректным для класса  $K$  набором ЭК.

Нетрудно видеть, что для корректного набора  $U$  всегда существует частичная булева функция, выполняющая роль корректирующей функции. Эта функция определена на откликах прецедентов и принимает значение 1, если прецедент из  $K$ , и принимает значение 0 в остальных случаях. Особо следует отметить корректные наборы ЭК, имеющие монотонную булеву функцию в качестве корректирующей функции и называемые монотонными. Очевидным является следующее утверждение.

**Утверждение 1.** Набор ЭК  $U = \{(\sigma_1, H_1), \dots, (\sigma_r, H_r)\}$  – монотонный корректный набор для класса  $K$  тогда и только тогда, когда для любых двух обучающих объектов  $S'$  и  $S''$  таких, что  $S' \in K$ ,  $S'' \notin K$ , можно указать  $i$  в  $\{1, 2, \dots, r\}$ , такое, что  $B_{(\sigma_i, H_i)}(S') = 1$  и  $B_{(\sigma_i, H_i)}(S'') = 0$ .

Требование монотонности можно убрать, если условие утверждения 1 заменить на условие  $B_{(\sigma_i, H_i)}(S') \neq B_{(\sigma_i, H_i)}(S'')$ .

Модели алгоритмов распознавания, основанные на построении корректных наборов ЭК, получили название логических корректоров.

При распознавании объекта  $S$  процедура голосования по корректному набору ЭК  $U$  заключается в следующем. Для каждой пары объектов  $(S, S')$ , где  $S' \in K$ , выписываются отклики  $U(S)$  и  $U(S')$  этих объектов на наборе  $U$ . Объект  $S$  не получает голос за принадлежность классу  $K$ , если не выполнено  $U(S') \preceq U(S)$  (здесь и далее запись  $U(S') \preceq U(S)$  означает, что каждая координата вектора  $U(S')$  не превосходит соответствующую координату вектора  $U(S)$ ). Если  $U$  – монотонный корректный набор ЭК класса  $K$  и  $U(S') \preceq U(S)$ , то объект  $S$  получает голос за принадлежность классу  $K$ . Если же  $U$  не является монотонным корректным набором ЭК класса  $K$ , то  $S$  получает голос за принадлежность к классу  $K$  только в случае  $U(S') = U(S)$ . Решение прикладных задач показало, что монотонный логический корректор работает лучше логического корректора с немонотонной корректирующей функцией.

На этапе построения семейств тупиковых корректных наборов ЭК, а также при построении процедур, основанных на голосовании по тупиковым представительным ЭК, приходится решать сложные дискретные задачи. Труднорешаемость этих задач обусловлена экспоненциальным ростом числа решений с ростом размера задачи и сложностью нахождения каждого нового решения. Одна из основных – монотонная дуализация. Задача допускает несколько формулировок. Ниже приведена ее матричная формулировка [13].

Пусть  $L$  – произвольная булева матрица. Набор различных столбцов  $W$  матрицы  $L$  называется покрытием, если каждая строка матрицы  $L$  в пересечении хотя бы с одним из столбцов,

входящих в  $W$ , даст 1 (т.е. в подматрице матрицы  $L$ , образованной столбцами из  $W$ , не должно быть строк вида  $(0, \dots, 0)$ ). Покрытие с минимальным числом столбцов называется минимальным. Покрытие называется неприводимым, если никакое его собственное подмножество не является покрытием. Неприводимость покрытия означает, что в подматрице матрицы  $L$ , образованной столбцами набора  $W$ , должна быть каждая из строк  $(1, 0, 0, \dots, 0, 0), (0, 1, 0, \dots, 0, 0), \dots, (0, 0, 0, \dots, 0, 1)$ . Требуется перечислить все неприводимые покрытия матрицы  $L$ . Алгоритмы перечисления неприводимых покрытий булевой матрицы применяются для поиска минимальных покрытий.

Другие формулировки монотонной дуализации используют понятия теории булевых функций (дуализация монотонной конъюнктивной нормальной формы) и понятия теории графов (дуализация гиперграфа) [14]. Задача имеет много важных приложений. Среди алгоритмов монотонной дуализации наиболее эффективны асимптотически оптимальные алгоритмы. Лидирующую позицию занимает алгоритм RUNC-M [13].

Как правило, даже для задач небольшой размерности процедура построения корректных наборов ЭК требует значительных вычислительных ресурсов. В простейших моделях логических корректоров для снижения вычислительной сложности используются ЭК ранга 1. При этом среди корректных наборов с ЭК, имеющими ранг, отличный от 1, могут оказаться наиболее информативные. В связи с этим для рассматриваемой задачи представляет интерес разработка не только точных, но и приближенных алгоритмов, например генетических и стохастических. Последние исследования в данной области посвящены также вопросам повышения качества решения прикладных задач на основе расширения класса корректирующих функций [4] и вопросам применения нейросетевых технологий [15].

Прикладные задачи классификации не всегда могут быть описаны в рамках классической постановки логической классификации, когда отдельные значения признака сравниваются с использованием простого отношения равенства. Довольно часто встречаются задачи классификации, в которых каждый признак принимает значения из некоторого частично упорядоченного множества. В работе [8] предложена схема синтеза корректных классических моделей логических алгоритмов распознавания при условии, что на множествах значений целочисленных признаков заданы конечные частичные порядки и признаковые описания изучаемых объектов – элементы декартона произведения заданных порядков. Фактически обобщены базовые понятия, используемые при логическом анализе целочисленных данных в задаче классификации по прецедентам, и приведены условия корректности основных логических процедур классификации. Аналогичным образом могут быть построены и логические корректоры над произведением частичных порядков. В условии содержания ЭК  $(\sigma, H)$ ,  $H = \{x_{j_1}, \dots, x_{j_r}\}$ ,  $\sigma = (\sigma_1, \dots, \sigma_r)$ , в объекте  $S$ ,  $S = (a_1, \dots, a_n)$ , следует заменить знак равенства на знак следования. Тогда соответствующее определение будет следующим: ЭК  $(\sigma, H)$  содержится в объекте  $S$ , если  $a_{j_i} \preceq \sigma_i$  при  $i = 1, r$ . Остальные введенные выше определения полностью сохраняются. Так же, как и в классическом случае (случае антицепей), для существования корректных наборов ЭК необходимо, чтобы описания объектов из разных классов были несравнимы.

**2. Построение монотонных логических корректоров с ЭК ранга 1.** Каждый (монотонный) корректный набор ЭК для класса  $K$  однозначно соответствует покрытию булевой матрицы  $L_K$ , построенной следующим образом по прецедентной информации.

Выпишем всевозможные ЭК, каждый из которых содержится в хотя бы одном обучающем объекте класса  $K$ . Пусть это будет множество ЭК  $\{Q_1, \dots, Q_u\}$ . Тогда каждой строке в  $L_K$  соответствует пара прецедентов  $S'$  и  $S''$  таких, что  $S' \in K$ ,  $S'' \notin K$ , а каждому столбцу соответствует один из ЭК  $Q_1, \dots, Q_u$ . На пересечении строки и столбца, порождаемых парой объектов  $(S', S'')$  и ЭК  $Q_j$ ,  $j = 1, u$ , ставится 1, если  $B_{Q_j}(S') = 1$  и  $B_{Q_j}(S'') = 0$ , т.е.  $Q_j$  содержится в  $S'$  и не содержитя в  $S''$ , иначе ставится 0. Нетрудно видеть, что набор ЭК является монотонным (тупиковым) корректным набором для  $K$  тогда и только тогда, когда соответствующий набор столбцов матрицы  $L_K$  является (неприводимым) покрытием.

Как правило, даже для задач небольшой размерности процедура перечисления монотонных тупиковых корректных наборов ЭК (МТК-наборов) для класса  $K$  с использованием матрицы  $L_K$  требует значительных вычислительных ресурсов, в связи с чем встает вопрос о разработке эффективных методов решения задач алгебрологической коррекции как в общем случае, так

и в следующих практически важных случаях: 1) каждый МТК-набор состоит из ЭК ранга 1; 2) в условиях задачи 1 на множестве  $N_j, j=1, n$ , допустимых значений признака  $x_j$  задан частичный порядок и считается, что ЭК  $(\sigma, \{x_j\})$  содержится в объекте  $S$ ,  $S=(a_1, \dots, a_n)$ , если  $a_j \preceq \sigma$ . В каждом из перечисленных случаев требуется разработка не только точных, но и приближенных методов решения задачи, что может оказаться необходимым для увеличения скорости ее решения.

Рассмотрим задачу перечисления МТК-наборов для класса  $K$ , порождаемых ЭК ранга 1 при условии, что каждое из множеств  $N_j, j=1, n$ , – антицепь, т.е. любые два элемента в  $N_j$  несравнимы. Пусть  $m_1$  – число прецедентов класса  $K$ ,  $m_2$  – число прецедентов не из класса  $K$ .

Нетрудно видеть, что в подматрице матрицы  $L_K = (a_{ij})$ ,  $i=1, m_1 m_2, j=1, 2n$ , образованной сравнением прецедента класса  $K$  с прецедентами из других классов, имеется максимум  $n$  ненулевых столбцов. Покажем, что указанное свойство матрицы  $L_K$  можно использовать для сокращения временных затрат при поиске МТК-наборов для класса  $K$  и искомый поиск может быть осуществлен на основе построения специальных наборов столбцов целочисленной матрицы с числом строк, равным числу строк матрицы  $L_K$  с числом столбцов, равным  $n$ .

Элементы  $a_{i_1 j_1}$  и  $a_{i_2 j_2}$  матрицы  $L_K$  назовем совместимыми, если  $i_1 \neq i_2, j_1 \neq j_2, a_{i_1 j_1} = a_{i_2 j_2} = 1$  и  $a_{i_1 j_1} = a_{i_2 j_2} = 0$ . Набор  $U$  из  $r$  единичных элементов матрицы  $L_K$  называется совместимым, если выполнено одно из следующих условий: 1)  $r = 1$ ; 2)  $r \geq 2$  и любые два элемента набора  $U$  совместимы.

Построим матрицу  $M_K = (b_{ij})$ ,  $i=1, m_1 m_2, j=1, 2n$ , образованную путем сравнения прецедентов класса  $K$  с прецедентами из других классов. Каждой строке этой матрицы соответствует пара прецедентов вида  $(S_i, S_u)$ , где  $S_i \in K, S_u \notin K$ . При этом если  $S_i$  содержит ЭК  $(\sigma, \{x_j\})$ , а  $S_u$  не содержит этот ЭК, то в строке, порожденной парой  $(S_i, S_u)$ , в столбце с номером  $j$  ставится  $\sigma + 1$ , иначе ставится 0.

Элементы  $b_{i_1 j_1}$  и  $b_{i_2 j_2}$  матрицы  $M_K$  называются совместимыми, если  $i_1 \neq i_2, b_{i_1 j_1} \neq 0, b_{i_2 j_2} \neq 0$  и дополнительно выполнено одно из двух следующих условий: 1)  $j_1 \neq j_2$  и  $b_{i_2 j_1} \neq b_{i_1 j_1}, b_{i_1 j_2} \neq b_{i_2 j_2}$ ; 2)  $j_1 = j_2, b_{i_1 j_1} \neq b_{i_2 j_2}$ . Набор  $V$  из  $r$  ненулевых элементов матрицы  $M_K$  называется совместимым, если выполнено одно из следующих условий: 1)  $r = 1$ ; 2)  $r \geq 2$  и любые два элемента набора  $V$  совместимы.

Заметим, что существует взаимно однозначное соответствие между ненулевыми элементами матриц  $L_K$  и  $M_K$ , при котором каждому единичному элементу  $e$  матрицы  $L_K$ , расположенному в строке с номером  $i$  и столбце, порожденном ЭК  $(\sigma, \{x_j\})$ , соответствует элемент  $\sigma + 1$  матрицы  $M_K$ , расположенный в строке с номером  $i$  и столбце с номером  $j$ , обозначаемый далее  $\Pi(e)$ .

**Утверждение 2.** Единичные элементы  $e_1$  и  $e_2$  совместимы в  $L_K$  тогда и только тогда, когда элементы  $\Pi(e_1)$  и  $\Pi(e_2)$  совместимы в  $M_K$ .

Доказательство. Пусть  $e_1$  и  $e_2$  – элементы матрицы  $L_K$ , такие, что  $e_1 = e_2 = 1$  и элемент  $e_t$ ,  $t \in (1, 2)$ , расположен в строке с номерами  $i_t$  и столбце, порожденном ЭК  $(\sigma_t, \{x_j\})$ . Обозначим через  $e_{t+2}$ ,  $t \in (1, 2)$ , элемент в  $L_K$ , расположенный в строке с номером  $i_t$  и в столбце, порожденном ЭК  $(\sigma, \{x_j\})$ , в котором  $\sigma = \sigma_2, j = j_2$  при  $t = 1$  и  $\sigma = \sigma_1, j = j_1$  при  $t = 2$ .

I. Пусть элементы  $e_1$  и  $e_2$  совместимы в матрице  $L_K$  и элемент  $e_t$ ,  $t \in (1, 2)$ , расположен в строке с номерами  $i_t$  и столбце, порожденном ЭК  $(\sigma, \{x_j\})$ . Предположим, что элементы  $\Pi(e_1)$  и  $\Pi(e_2)$  не совместимы в  $M_K$ . Тогда в случае  $j_1 \neq j_2$  получаем, что хотя бы один из элементов  $e_3$  и  $e_4$  не равен 0. Если же  $j_1 = j_2$ , то  $\sigma_1 = \sigma_2$ , т.е. элементы  $e_1$  и  $e_2$  расположены в одном и том же столбце матрицы  $L_K$ . Таким образом, в обоих случаях возникает противоречие с тем, что  $e_1$  и  $e_2$  совместимые в  $L_K$  элементы.

II. Пусть элементы  $\Pi(e_1)$ ,  $\Pi(e_1) = \sigma_1 + 1$ , и  $\Pi(e_2)$ ,  $\Pi(e_2) = \sigma_2 + 1$ , совместимы в матрице  $M_K$  и расположены в строках  $i_1$  и  $i_2$  соответственно. Тогда  $e_1 = e_2 = 1$ .

Предположим, что элементы  $e_1$  и  $e_2$  не совместимы в  $L_K$  и  $j_1 \neq j_2$ . Это означает, что выполнено хотя бы одно из двух равенств  $e_3 = 1$  и  $e_4 = 1$ , где  $e_{t+2}$ ,  $t \in \{1, 2\}$ , – элемент в  $L_K$ , расположенный в строке с номером  $i_t$  и в столбце, порожденном ЭК  $(\sigma, \{x_j\})$ , в котором  $\sigma = \sigma_2, j = j_2$  при  $t = 1$  и  $\sigma = \sigma_1, j = j_1$  при  $t = 2$ . Если, например,  $e_3 = 1$ , то  $\Pi(e_3) = \Pi(e_2) = \sigma_2 + 1$ . Противоречие с условием совместимости элементов  $\Pi(e_1)$  и  $\Pi(e_2)$ .

Допустим, что элементы  $e_1$  и  $e_2$  не совместимы в  $L_K$  и  $j_1 = j_2 = j$ . Тогда существует  $t \in (1, 2)$ , такое, что на пересечении строки с номером  $i_t$  матрицы  $L_K$  со столбцами, порождаемыми ЭК

$(\sigma_1, \{x_j\})$  и  $(\sigma_2, \{x_j\})$ , стоит число 1, что недопустимо при  $\sigma_1 \neq \sigma_2$ . Следовательно,  $\sigma_1 = \sigma_2$ . Противоречие с условием совместимости элементов  $\Pi(e_1)$  и  $\Pi(e_2)$ .

Утверждение 2 доказано.

Пусть  $\{\sigma_1, \dots, \sigma_r\}$  – совместимый набор элементов матрицы  $M_K$ , в котором  $\sigma_i$  – элемент столбца с номером  $j_i$ , и пусть  $H$  – набор столбцов с номерами  $j_1, \dots, j_r$ . Через  $\tilde{H}$  обозначим набор из всех различных столбцов, встречающихся в  $H$ . Будем говорить, что  $H$  –  $(\sigma_1, \dots, \sigma_r)$  ( $\sigma_1, \dots, \sigma_r$ ) – совместимый набор столбцов. Будем считать, что для  $H$  выполнено условие покрываемости, если подматрица  $M_K^H$  матрицы  $M_K$ , образованная столбцами из  $\tilde{H}$ , не содержит строк вида  $(0, \dots, 0)$ . Назовем  $H$  корректным ( $\sigma_1, \dots, \sigma_r$ ) покрытием для  $M_K$ , если для  $H$  выполнено условие покрываемости. Корректное покрытие матрицы  $M_K$ , имеющее наименьшее число столбцов, назовем минимальным. Согласно утверждению 2, существует взаимно однозначное соответствие между множеством корректных покрытий матрицы  $M_K$  и множеством неприводимых покрытий матрицы  $L_K$ . Следовательно, для синтеза МТК-наборов класса  $K$  могут быть применены алгоритмы построения корректных покрытий матрицы  $M_K$ , а для синтеза минимальных по мощности МТК-наборов класса  $K$  могут быть применены алгоритмы построения минимальных корректных покрытий матрицы  $M_K$ .

Для перечисления корректных покрытий матрицы  $M_K$  авторами построен алгоритм АЛКОР. Ниже приведена схема работы этого алгоритма.

Предположим, что  $C(M_K)$  – множество всех совместимых наборов столбцов матрицы  $M_K$  и пусть  $H \in C(M_K)$ . Столбец  $h$  матрицы  $M_K$  совместим с  $H$ , если набор столбцов  $H \cup h$  принадлежит  $C(M_K)$ , иначе  $h$  несовместим с  $H$ . Если в матрице  $M_K$  не существует столбца, совместимого с  $H$ , то  $H$  – максимальный совместимый набор столбцов этой матрицы. Допустим, что  $C^*(M_K)$  – множество всех максимальных совместимых наборов столбцов матрицы  $M_K$ .

Для наглядности работу алгоритма АЛКОР можно представить в виде обхода дерева решений (ДР) в глубину. Корень ДР – пустой набор, вершина ДР – набор из  $C(M_K)$ , висячая вершина ДР – набор из  $C^*(M_K)$ , которая либо являются корректными покрытием матрицы  $M_K$ , найденным впервые, либо соответствует лишнему шагу.

Первоначально матрица  $M_K$  рассматривается в качестве текущей матрицы. Построение ДР начинается с вершины, которая порождается некоторым ненулевым столбцом с номером  $j_1$  матрицы  $M_K$  и некоторым ненулевым элементом  $\sigma_1$  этого столбца. Далее в роли текущей матрицы выступает подматрица матрицы  $M_K$ , полученная описанным ниже способом. Для построения дочерней вершины для вершины  $(\sigma_1, j_1)$  алгоритм по определенному правилу выбирает в текущей матрице ненулевой столбец с номером  $j_2$  и элемент  $\sigma_2$  этого столбца, совместимый с элементом  $\sigma_1$ . Строится новая вершина  $(\sigma_2, j_2)$ . При этом возможно, что  $j_1 = j_2$ . После построения очередной вершины  $(\sigma_r, j_r)$ ,  $r \geq 1$ , порождаемой  $(\sigma_1, \dots, \sigma_r)$ -совместимым набором столбцов  $H$  матрицы  $M_K$ , текущая матрица изменяется следующим образом. Из нее удаляются все строки, дающие  $\sigma_r$  в пересечении со столбцом с номером  $j_r$ .

Алгоритм АЛКОР применяет “жадную” стратегию выбора столбцов для построения дочерних вершин в ДР. Всякий раз после построения очередной вершины  $(\sigma_r, j_r)$ ,  $r \geq 1$ , в новой текущей матрице среди строк с наименьшим числом ненулевых элементов выбирается строка с наименьшим номером. Далее среди столбцов, которые в пересечении с выбранной строкой дают отличные от 0 элементы, ищется столбец с наименьшим номером  $j_{r+1}$ , такой, что ненулевой элемент  $\sigma_{r+1}$  этого столбца, расположенный в выбранной строке, совместим с текущим совместимым набором  $\{\sigma_1, \dots, \sigma_r\}$  элементов матрицы  $M_K$ . Если указанный столбец найти не удается, то вершина  $(\sigma_r, j_r)$  становится висячей. Далее либо происходит переход к новому шагу путем возврата на более высокий уровень ДР, либо алгоритм заканчивает работу. В противном случае  $(\sigma_r, j_r)$  – внутренняя вершина ДР, процесс построения текущей ветви дерева продолжается, и строится новая вершина  $(\sigma_{r+1}, j_{r+1})$ . Если можно выбрать только один столбец, обладающий названными выше свойствами, то  $(\sigma_{r+1}, j_{r+1})$  – концевая вершина.

Пусть висячая вершина  $(\sigma_r, j_r)$  порождает максимальный  $(\sigma_1, \dots, \sigma_r)$ -совместимый набор столбцов  $H$  матрицы  $M_K$ . В этом случае для  $H$  проверяется условие покрываемости. Если условие покрываемости для  $H$  выполнено, то множество корректных покрытий матрицы  $M_K$ , построенных на предыдущих шагах, пополняется добавлением к нему набора столбцов  $H$ , в противном случае это множество не меняется. Алгоритм заканчивает работу, если каждая из вершин  $(\sigma_i, j_i)$ ,  $i = 1, r$ , является концевой вершиной.

Таблица 1. Эффективность алгоритмов RUNC-M+ и АЛКОР+ на модельных данных

Номер задачи	Размер задачи		RUNC-M+	АЛКОР+
	$m \times n$			
1	$20 \times 40$	76.2	0.9	0.4
2	$20 \times 60$	104.2	6.7	2.6
3	$20 \times 80$	169	24.2	3.6
4	$20 \times 100$	269.4	96.9	10.6
5	$20 \times 120$	861.3	115.1	17.7
6	$20 \times 140$	397	150.9	20.9
7	$20 \times 160$	422.6	137.1	24.7
8	$20 \times 180$	642.8	256.4	42.3
9	$20 \times 200$	1131.3	409.3	61.3

Таким образом, число шагов алгоритма АЛКОР равно числу висячих вершин в ДР. Сложность шага алгоритма  $O(qm_1m_2n)$ ,  $q = \min(m_1, n)$ .

Схема работы алгоритма АЛКОР близка к схеме работы асимптотически оптимального алгоритма перечисления неприводимых покрытий булевой матрицы RUNC-M [13]. В настоящей работе на базе алгоритмов RUNC-M и АЛКОР построены алгоритмы RUNC-M+ и АЛКОР+, осуществляющие соответственно поиск минимальных покрытий матрицы  $L_K$  и минимальных корректных покрытий матрицы  $M_K$ . В этих алгоритмах добавлена текущая переменная, фиксирующая минимальное число столбцов в найденных покрытиях.

В случае, когда на каждом множестве  $N_j$ ,  $j = 1, n$ , допустимых значений признака  $x_j$  задан частичный порядок, применялся аналогичный подход. Проведены эксперименты, показавшие эффективность предложенной методики решения поставленных задач. В экспериментах при решении реальных задач рассматривался специальный линейный порядок, учитывающий частоту встречаемости значения признака в описаниях прецедентов [16].

**3. Результаты экспериментов.** Выполнены эксперименты на случайных модельных бинарных данных с равновероятным появлением 0 и 1, с равномощными непересекающимися классами  $K_1$  и  $K_2$ , с числом прецедентов  $m$  и числом признаков  $n$ . В табл. 1 приведено среднее время (по 20 запускам) поиска минимальных покрытий матрицы  $L_K$ ,  $K \in \{K_1, K_2\}$ , алгоритмами RUNC-M+ и АЛКОР+. Время счета указано в секундах. Дополнительно указано среднее число минимальных покрытий  $q$ .

Из результатов, представленных в табл. 1, следует, что алгоритм АЛКОР+ работает существенно быстрее алгоритма RUNC-M+ в случае, когда  $m \ll n$ . В среднем при  $n \geq 100$  время счета алгоритма RUNC-M+ в 6.5 раз больше, а при  $n < 100$  – в 3.5 раза. Проведены также аналогичные эксперименты с фиксированным числом признаков, равным 20, и с числом прецедентов, меняющимся от 40 до 200, которые показали, что по времени счета АЛКОР+ незначительно уступает RUNC-M+ (до 1.5 раз).

Выполнены эксперименты на пяти реальных задачах с двумя классами, содержащими как бинарные признаки, так и целочисленные признаки высокой значности. Одна задача (“tic-tac-toe”) была взята из репозитория UCI [17], а остальные – из репозитория ФИЦ ИУ РАН [11]. Тестирование новой модели монотонного логического корректора (МКОР), проводилось с линейными порядками на множествах значений признаков, согласно частоте встречаемости каждого значения [16]. Исходная обучающая выборка 10 раз разбивалась на обучение и контроль в соотношении 80 и 20%. Качество классификации оценивалось двумя функционалами:  $Q = (\tilde{m}_1 + \tilde{m}_2) / (m_1 + m_2)$  – средняя точность и  $Q_b = 1 / 2 (\tilde{m}_1 / m_1 + \tilde{m}_2 / m_2)$  – сбалансированная точность, где  $m_i$ ,  $i \in \{1, 2\}$ , – число контрольных объектов класса  $K_i$ , а  $\tilde{m}_i$  – число верно классифицированных контрольных объектов класса  $K_i$ . Для задачи “tic-tac-toe” указаны результаты счета при поиске всех неприводимых покрытий в  $L_K$ , так как в  $L_K$  всегда было найдено только одно минимальное покрытие. Результаты счета представлены в табл. 2, в которой указаны также среднее число найденных покрытий  $q$  и максимальная значность признака  $k$ .

Таблица 2. Эффективность алгоритма МКОР на реальных данных

Описание задачи		$q$	$Q/Q_b$	RUNC-M+	АЛКОР+
Название	$m \times n (k)$				
insult	$79 \times 81 (2)$	351.2	0.69/0.55	10467.8	1361.6
tic-tac-toe	$958 \times 9 (3)$	112.8*	0.99/0.99	7.4	5.7
dorovski	$33 \times 12 (33)$	1297.8	0.60/0.60	2086.3	351.9
input	$344 \times 9 (11)$	1220.2	0.95/0.94	60989.4	14538.7
sigapur	$58 \times 15 (58)$	34	0.98/0.99	1.1	0.05

Таблица 3. Показатели качества тестируемых классификаторов

Название задачи	МКОР $Q/Q_b$	RF $Q/Q_b$	LADTree $Q/Q_b$	NB $Q/Q_b$	ГТП $Q/Q_b$	ГТП+ $Q/Q_b$
insult	0.69/0.55	<b>0.79/0.55</b>	0.64/0.49	0.76/0.55	0.68/0.50	<b>0.76/0.58</b>
tic-tac-toe	<b>0.99/0.99</b>	0.95/0.93	0.76/0.70	0.64/0.71	0.69/0.56	0.69/0.56
dorovski	0.60/0.60	0.50/0.50	0.63/0.66	0.57/0.56	0.53/0.53	<b>0.69/0.69</b>
input	0.95/0.94	0.97/0.97	0.96/0.96	<b>0.99/0.99</b>	0.64/0.50	0.64/0.50
sigapur	<b>0.98/0.99</b>	0.97/0.93	0.94/0.94	0.94/0.91	0.82/0.50	0.82/0.50

Из полученных результатов следует, что алгоритм АЛКОР+ работает быстрее алгоритма RUNC-M+, причем существенно быстрее на задачах с большим числом признаков. Время счета в секундах указано в последних двух столбцах табл. 2.

В табл. 3 приведены результаты сравнения качества алгоритма МКОР с такими известными алгоритмами, как Random Forest (RF), бустинг над решающими деревьями (LADTree), наивный байес (NB), алгоритм голосования по всем тупиковым представительным ЭК (ГТП) и модификация алгоритма ГТП, в которой “голосуют” тупиковые представительные ЭК с небольшим рангом (ГТП+). При счете использовалась авторская реализация алгоритмов ГТП и ГТП+. Алгоритмы RF, LADTree и NB были взяты из системы WEKA [18].

Согласно результатам, представленным в табл. 3, качество алгоритма МКОР сопоставимо с другими тестируемыми классификаторами, а на задачах “tic-tac-toe” и “sigapur” МКОР лидирует. При этом качество МКОР на большинстве задач выше качества ГТП, особенно в случае задач с высокой значностью признаков.

На задаче “insult” проведено исследование стохастического варианта классификатора МКОР (табл. 4). Для каждого из 10 случайных разбиений исходной выборки на обучение и контроль в соотношении 80 и 20% на вход  $w$  раз подавался случайный набор из  $p$  различных столбцов матрицы  $M_K$ . Значение параметра  $w$  варьировалось от 20 до 200 с шагом 20 и рассматривалось два значения параметра  $p$ , а именно  $p_1 = 0.2n$  и  $p_2 = 0.25n$ . Для каждого  $p$  и  $w$  учитывались все найденные минимальные покрытия. В табл. 4 также указано среднее время счета  $t$  в миллисекундах и среднее число покрытий  $q$  по 10 разбиениям исходной выборки.

Таблица 4. Эффективность стохастического варианта алгоритма МКОР на задаче “insult”

$w$	$p = p_1$			$p = p_2$		
	$Q/Q_{balance}$	$t$ , мс	$q$	$Q/Q_{balance}$	$t$ , мс	$q$
20	0.76/0.51	18.1	56	0.80/0.51	230.3	250.2
40	0.81/0.54	17.4	162.7	0.81/0.51	230.9	692.3
60	0.74/0.51	14.3	53.3	0.78/0.48	338.8	588.9
80	0.78/0.53	11.2	131.1	0.79/0.49	465.5	682.5
100	0.78/0.52	22.4	186.8	0.81/0.49	582.2	1248.5
120	0.78/0.53	17.2	126.5	0.78/0.48	1377.7	1159.1
140	0.78/0.49	26.4	245.2	0.79/0.49	527.2	1378.8
160	0.79/0.51	62.5	244.9	0.81/0.51	1075.9	1612.3
180	0.79/0.48	33.9	267.4	0.81/0.50	896.4	2553.5
200	0.790/0.50	42.6	237.7	0.81/0.50	1459.9	2454.9

Нетрудно видеть, что при  $p = p_1$  стохастический вариант алгоритма МКОР показывает схожие результаты с детерминированным вариантом по функционалу  $Q_b$  и лучшие результаты по функционалу  $Q$ . Время стохастического поиска существенно меньше (в среднем в 40 тыс. раз). При увеличении  $w$  качество стохастического алгоритма немного снижается, при этом увеличивается время счета (в среднем в 3 раза). При увеличении параметра  $p$  качество стохастического алгоритма немного снижается по функционалу  $Q_b$ , но существенно увеличивается время счета (в среднем в 25 раз).

Таким образом, проведенные эксперименты свидетельствуют о целесообразности применения алгоритма АЛКОР+ для построения монотонного логического корректора с ЭК ранга 1. Установлено, что снижение временных затрат особенно заметно, когда число признаков существенно больше числа обучающих объектов. Показано, что предлагаемый стохастический классификатор не уступает по качеству своему детерминированному варианту и работает на реальных задачах в тысячи раз быстрее.

**Заключение.** Рассмотрены вопросы повышения эффективности логического подхода к решению задачи корректной классификации по прецедентам. Разработаны и экспериментально исследованы новые более совершенные детерминированные и стохастические модели монотонных логических корректоров, предназначенные в том числе на обработку частично упорядоченных данных и позволяющие существенно сократить временные затраты этапа обучения.

## СПИСОК ЛИТЕРАТУРЫ

- Чегис И.А., Яблонский С.В. Логические способы контроля электрических схем // Сб. статей по математической логике и ее приложениям к некоторым вопросам кибернетики. Тр. МИАН СССР. М.: АН СССР, 1958. Т. 51. С. 270–360.
- Дюкова Е.В., Журавлëв Ю.И., Рудаков К.В. Об алгебро-логическом синтезе корректных процедур распознавания на базе элементарных алгоритмов // ЖВМ и МФ. 1996. Т. 36. № 8. С. 217–225.
- Djukova E.V., Zhuravlev Yu.I., Sotnezhov R.M. Construction of an Ensemble of Logical Correctors on the Basis of Elementary Classifiers // Pattern Recognition and Image Analysis. 2011. V. 21. № 4. P. 599–605.
- Дюкова Е.В., Журавлëв Ю.И., Прокофьев П.А. Логические корректоры в задаче классификации по прецедентам // ЖВМ и МФ. 2017. Т. 57. № 11. С. 1906–1927.
- Журавлëв Ю.И. Об алгебраическом подходе к решению задач распознавания или классификации // Пробл. кибернетики. М.: Наука, 1978. Вып. 33. С. 5–68.
- Воронцов К.В. Оптимизационные методы линейной и монотонной коррекции в алгебраическом подходе к проблеме распознавания // ЖВМ и МФ. 2000. Т. 40. № 1. С. 166–176.
- Абламайко С.В., Бирюков А.С., Докукин А.А., Дьяконов А.Г., Журавлев Ю.И., Краснопрошин В.В., Образцов В.А., Романов М.Ю., Рязанов В.В. Практические алгоритмы алгебраической и логической коррекции в задачах распознавания по прецедентам // ЖВМ и МФ. 2014. Т. 54. № 12. С. 1979–1993.
- Дюкова Е.В., Масляков Г.О., Прокофьев П.А. О логическом анализе данных с частичными порядками в задаче классификации по прецедентам // ЖВМ и МФ. 2019. Т. 59. № 9. С. 1605–1616.
- Баскакова Л.В., Журавлëв Ю.И. Модель распознающих алгоритмов с представительными наборами и системами опорных множеств // ЖВМ и МФ. 1981. Т. 21. № 5. С. 1264–1275.
- Дюкова Е.В., Масляков Г.О., Дюкова А.П. Логические методы корректной классификации данных // Информатика и ее применения. 2023. Т. 17. Вып. 3. С. 64–70.
- Журавлëв Ю.И., Рязанов В.В., Сенько О.В. Распознавание. Математические методы. Программная система. Практические применения. М.: ФАЗИС, 2006. 159 с.
- Дюкова Е.В., Сизов А.В., Сотников Р.М. Об оптимальном корректном перекодировании целочисленных данных в распознавании // Информатика и ее применения. 2012. Т. 6. Вып. 4. С. 61–65.
- Дюкова Е.В., Прокофьев П.А. Об асимптотически оптимальных алгоритмах дуализации // ЖВМ и МФ. 2015. Т. 55. № 5. С. 895–910.
- Хачиян Л.Г. Избранные труды [сост. С. П. Тарасов]. М.: МЦНМО, 2009. 520 с.
- Лютикова Л.А., Шматова Е.В. Логический подход к коррекции результатов работы SP-нейронных сетей // Информационные технологии. 2018. Т. 24. № 2. С. 110–116.
- Дюкова Е.В., Масляков Г.О., Янаков Д.С. Корректная классификация по прецедентам: ДСМ-метод над произведением частичных порядков // Информатика и ее применения. 2024. Т. 18. Вып. 3. С. 61–68.
- Asuncion A., Newman D. 2007. UCI Machine Learning Repository. <https://archive.ics.uci.edu/>
- WEKA: Suite of Machine Learning Software, Developed at the University of Waikato. New Zealand, 2017. <http://www.cs.waikato.ac.nz/~ml/weka/>