

17. *Klarens Kho, Rob Kharrop*. Spring 3 dlya professionalov = Pro Spring 3 [Spring 3 for pros = Pro Spring 3]. Moscow: Vil'yams, 2012, 880 p.
18. *Kartiev S.B., Kureychik V.M.* Algoritm klassifikatsii, osnovannyi na printsipakh sluchainogo lesa dlya resheniya zadachi prognozirovaniya [The classification algorithm is based on the principles of random forests for forecasting], *Programmnye produkty i sistemy* [Software Products and Systems], 2016, No. 2, pp. 11-15.
19. *Kartiev S.B., Kureychik V.M. Martynov A.V.* Parallelnyy algoritm prognozirovaniya korotkikh vremennykh ryadov [A parallel algorithm for forecasting short time series], *Trudy Kongressa po intellektual'nyim sistemam i in-formatsionnym tekhnologiyam «IS&IT'15»*. Nauchnoe izdanie v 4-kh t. [Proceedings of Congress on intelligent systems and information technologies "IS&IT'15". Scientific publication in 4 vol.]. Moscow: Fizmatlit, 2015, pp. 27-47.
20. *Kartiev S.B., Kureychik V.M.* Razrabotka raspredelennoy sistemy analiza vremennykh ryadov na osnove modeli vychisleniya MapReduce [Development of a distributed system for analyzing time series based on the model of MapReduce computation], *Trudy Kongressa po intellektual'nyim sistemam i informatsionnym tekhnologiyam «IS&IT'16»*. Nauchnoe izdanie v 4-kh t. [Proceedings of Congress on intelligent systems and information technologies "IS&IT'16". Scientific publication in 4 vol.]. Moscow: Fizmatlit, 2016, pp. 36-43.

Статью рекомендовал к опубликованию д.т.н., профессор Я.Е. Ромм.

Картиев Санчир Басанович – Южный федеральный университет; e-mail: mlearningsystems@gmail.com; 347928, г. Таганрог, пер. Некрасовский, 44; кафедра ДМиМО; аспирант.

Курейчик Виктор Михайлович – e-mail: kur@tsure.ru; кафедра ДМиМО; профессор.

Kartiev Sanchir Basanovich – South Federal University; e-mail: mlearningsystems@gmail.com; 44, Nekrasovskiy, Taganrog, 347928, Russia; the department of discrete mathematics and optimization methods; postgraduate student.

Kureychik Viktor Michailovich – e-mail: kur@tsure.ru; the department of discrete mathematics and optimization methods; professor.

УДК 681.3.06:378.1

DOI 10.18522/2311-3103-2016-7-2839

И.И. Казмина, Е.В. Нужнов

МОДИФИКАЦИЯ АПРИОРНОГО АЛГОРИТМА ДЛЯ АНАЛИЗА ДАННЫХ УЧЕБНОГО ПРОЦЕССА ВУЗА*

Интеллектуальный анализ данных (ИАД) учебного процесса является одним из механизмов, позволяющих получить больше полезных сведений из имеющихся массивов данных и использовать полученные результаты с целью повышения эффективности и качества образовательной деятельности. Преимуществом использования ИАД является возможность выявления скрытых закономерностей, которые не всегда видны при использовании статистических методов. Априорный алгоритм ИАД позволяет на основе анализа больших массивов исходных данных выявлять зависимости между часто встречающимися элементами данных и анализируемой величиной. В качестве анализируемой величины в данном случае выступает успеваемость студентов, которая численно отражает эффективность учебного процесса, а исходными являются различные данные, касающиеся образовательной деятельности. Априорный алгоритм может выявить большое число правил в исходных данных, значительная часть которых может быть заранее известна пользователю, вследствие чего неинформативна. Для устранения этой проблемы в работе предлагается модификация Априорного алгоритма, учитывающая такой показатель правил, как их инфор-

* Исследование выполнено за счет гранта Российского научного фонда (проект № 14-11-00242) в Южном федеральном университете.

мативность. Данный показатель рассчитывается на основании оценки пользователем важности рассмотрения тех или иных элементов данных. Пользователь может изменять данные оценки в зависимости от целей анализа. Таким образом, показатель информативности всегда будет иметь актуальные значения, а выводимые правила будут информативны для пользователя. Предложенная модификация алгоритма позволит выявлять и предоставлять пользователю, в первую очередь, наиболее актуальные для него зависимости в исходных данных. Выявленные взаимосвязи между успеваемостью и данными об учебном процессе можно использовать для его корректировки с целью повышения эффективности и качества.

Интеллектуальный анализ данных; Априорный алгоритм; образовательная деятельность; подсистема ИАД.

I.I. Kazmina, E.V. Nuzhnov

MODIFICATION OF A PRIORI ALGORITHM FOR ANALYSIS UNIVERSITY EDUCATIONAL PROCESS DATA

Data mining of educational process data is one of mechanisms which allow to get more useful information of existing data array and to use derived results with the view of increase of educational activities effectiveness and quality. Advantage of using of data mining algorithms is possibility of exposure of hidden laws which are visible no always by use statistical technology. A priori algorithm enable reveal of interrelation between abundant data elements and analyzable value by virtue of analysis of a big array of raw data. In this case analyzable value is progress in studies which numerically reflect educational process effectiveness. Raw data are different information about educational activities. The A priori algorithm is able to reveal a big number of rules in raw data, considerable part of which can be known to user beforehand, therefore are non-informative. For the purpose of removal this problem in this paper the modification of A priori algorithm is suggested, which take account of rules factor as information value. This factor is calculated on the basis of assessment of data elements interest by user. User can change this assessment depend on purpose of analysis. So information value factor always will have topical value, and deducible rules will informative for user. Suggested modification of A priori algorithm enables to reveal and give user in the first place the most topical for him relations into raw data. Revealed relations between progress in studies and educational process data can use for educational process adjusting with the purpose of its efficiency and quality improvement.

Data mining; a priori algorithm; educational activities; data mining subsystem.

Введение. Одна из основных задач улучшения качества образования – повышение уровня знаний, умений и навыков обучаемых, что численно выражается в виде успеваемости. Для повышения уровня успеваемости целесообразно выявить влияющие на него аспекты и установить характер такого влияния. Тогда зная, что приводит к улучшению успеваемости, а что – к ее понижению, можно внести такие изменения в учебный процесс, которые заведомо улучшат его качество. Некоторые аспекты, влияющие на успеваемость вполне понятны и видны при экспертном рассмотрении проблемы, но некоторые из них могут быть не видны на первый взгляд, но оказывать существенное влияние на эффективность обучения. Такие, невидимые на первый взгляд факторы можно выявить путем ИАД учебного процесса.

ИАД предоставляет эффективные инструменты для анализа информации различной природы [1–4]. Для этого используются разнообразные алгоритмы ИАД, каждый из которых решает свои задачи [5–7]. Так, широко известны алгоритмы кластеризации, выявления взаимосвязей, регрессии, дерева зависимостей и др. [8]. Интеллектуальная обработка и анализ информации включают набор методов эвристической и автоматизированной обработки данных [9]. В рамках рассматриваемой в работе проблемы наиболее эффективным представляется алгоритм выявления взаимосвязей – Априорный алгоритм.

Априорный алгоритм выявления взаимосвязей был предложен в 1994 г. R. Agrawal и R. Srikant [10]. В основу этого алгоритма заложено использование априорных данных о частоте выбора покупателями тех или иных наборов продуктов [11].

В настоящее время Априорный алгоритм успешно используется в системах ИАД. Так, Алгоритм правил взаимосвязей (Майкрософт) представляет собой простую реализацию априорного алгоритма [12]. Априорный алгоритм используется в системе Oracle Data Mining [13]. В пакете SAP Predictive Analysis реализованы Априорный алгоритм HANA, Априорный алгоритм (R) и Априорный алгоритм HANA (R) [14], различающиеся набором задаваемых пользователем параметров.

Априорный алгоритм может использоваться в различных предметных областях, таких как анализ потребительской корзины, прогнозирование курсов валют и др. Во многих сферах данный алгоритм дает хорошие результаты, в связи с чем целесообразно рассмотреть возможности его применения в образовательной деятельности.

При работе с большими объемами данных априорный алгоритм обычно дает хорошие результаты, но в тех случаях, когда объем данных невелик, получаемые результаты могут оказаться плохо объяснимыми с точки зрения здравого смысла, а иногда даже являются ложными [11]. Учебные заведения, как правило, накапливают большие объемы информации, относящиеся к учебному процессу, в связи с чем, рассматриваемый алгоритм должен давать хорошие результаты.

При анализе потребительской корзины, прогнозировании курса валют и решении других подобных задач данные для анализа, как правило, являются однородными: в первом случае это наименования товаров, во втором – показатели курса и т.д. В образовательной деятельности данные могут быть весьма разнородными – это и используемые учебно-методические материалы, и число проведенных лекций, и публикационная активность студентов, и другая информация. В связи с этим целесообразно учитывать особенности предметной области при анализе данных.

В данной работе рассмотрен Априорный алгоритм ИАД и предложена его модификация для учебного процесса, учитывающая разнородность исходных данных и информационные потребности пользователя.

1. Априорный алгоритм ИАД. Априорный алгоритм предназначен для поиска повторяющихся наборов элементов и выявления на их основе закономерностей (взаимосвязей) в исходных данных. Под элементом понимают отдельный элемент данных, в реляционных базах данных – информация, записанная в ячейку таблицы. Набор – это несколько элементов, встречающихся вместе (в пределах одной строки) в таблице исходных данных. Набор может состоять из любого числа элементов. Для характеристики (анализа) выявленных наборов используют два основных показателя: **поддержка** (support) и **достоверность** (confidence). Эти показатели отражают значимость наборов элементов и используются для отсеивания менее важных из них. Для отсеивания задают пороговые значения показателей поддержки и достоверности. Таким образом, отсеиваются наборы, величины поддержки и достоверности для которых ниже соответствующих пороговых значений.

Обозначим некоторый набор элементов A , его поддержку – $S(A)$, тогда поддержку набора элементов можно рассчитать по формуле:

$$S(A) = N_s(A)/N,$$

где $N_s(A)$ – число строк в таблице, в которых встречается набор A ; N – общее число строк в таблице. Поддержка характеризует частоту встречаемости набора элементов в общем объеме данных.

Достоверность характеризует частоту возникновения события, когда присутствие набора элемента привело к наличию определенного состояния прогнозируемой величины. Таким образом, достоверность – это характеристика правила. Обозначим $C(A \rightarrow B)$ достоверность правила вида $A \rightarrow B$. Достоверность можно рассчитать по формуле:

$$C(A \rightarrow B) = S(A \rightarrow B) / S(A).$$

Правила, которые имеют высокие значения показателей поддержки и достоверности, называют сильными, и именно они представляют наибольший интерес при выявлении и анализе закономерностей в исходных массивах данных.

Полный перебор всех возможных вариантов при больших объемах исходных данных становится трудоемким, поэтому для сокращения объема вычислений используют пороговые значения, позволяющие отсеивать наименее важные наборы элементов и правила [11–12].

После выявления наборов элементов Априорный алгоритм оставляет для дальнейшего рассмотрения только те из них, величина поддержки которых больше некоторого заранее заданного порогового значения, т.е. те, для которых выполняется условие:

$$S(A) \geq S_{min}, \quad (1)$$

где S_{min} – пороговое значение величины поддержки наборов элементов.

При этом сначала рассматриваются наборы из одного элемента, и удаляются те из них, которые имеют поддержку ниже порогового значения. Далее рассматриваются наборы из двух элементов, которые, в свою очередь, прошли предыдущий шаг отбора и т.д. Т.е. на каждом i -м шаге на предмет выявления величины поддержки рассматриваются наборы, состоящие из других наборов, меньшей размерности, прошедших отбор на предыдущем ($i-1$)-ом шаге. И так продолжается до тех пор, пока не будут рассмотрены все наборы наибольшей найденной размерности. В результате такой проверки для дальнейшего рассмотрения остаются только наборы элементов, все подмножества которых удовлетворяют условию (1). Получившиеся наборы элементов обладают антимонотонным свойством поддержки [15], имеющим в соответствии с введенными выше обозначениями следующий вид:

$$\forall A, B: (B \subseteq A) \rightarrow S(B) \geq S(A),$$

где A, B – наборы элементов; $S(A), S(B)$ – величины поддержки соответствующих наборов элементов.

Из наборов элементов, оставшихся после отсева по условию (1), строятся правила вида $A \rightarrow B$. Далее, из числа этих правил отсеиваются те, величина достоверности которых меньше некоторого заранее заданного значения, т.е. те, для которых не выполняется условие:

$$C(A \rightarrow B) \geq C_{min}, \quad (2)$$

где C_{min} – пороговое значение величины достоверности правил.

Правила, прошедшие проверку по условию (2), являются искомыми ассоциативными правилами, на основании которых можно судить о взаимосвязях в исходных данных и прогнозировать значение интересующего элемента данных в зависимости от значений других.

При задании пороговых значений поддержки и достоверности следует помнить, что увеличение значений этих величин приводит к уменьшению числа итераций алгоритма. Это связано с тем, что на каждом шаге будет отсеиваться большее число наборов элементов, следовательно, будет оставаться меньше данных для дальнейшего рассмотрения.

Кроме того, для выявленных правил может быть рассчитан еще один дополнительный показатель – **подъем** (lift). Данный показатель характеризует наличие обратной зависимости между исходом и посылкой правила. Подъем для правила вида $A \rightarrow B$ обозначим $L(A \rightarrow B)$. Данную величину можно рассчитать следующим образом:

$$L(A \rightarrow B) = C(A \rightarrow B) / S(B).$$

Показатель подъема обладает следующим свойством:

$$L(A \rightarrow B) = L(B \rightarrow A).$$

На основании описания Априорного алгоритма можно представить его схему. Для этого введем следующие обозначения:

- ◆ M – множество наборов элементов, найденных в исходных данных. $M = \{M^1, M^2, \dots, M^I\}$, где: M^i – множество наборов элементов размерности i ; $i = 1, 2, \dots, I$; I – максимальная размерность наборов элементов;
- ◆ $M^i = \{M^i_1, M^i_2, \dots, M^i_J\}$, где M^i_j – j -й набор элементов размерности i , J – общее число найденных наборов элементов размерности i .
- ◆ R – множество правил вида $A \rightarrow B$, полученных из M . $R = \{R_1, R_2, \dots, R_K\}$, где K – число найденных правил.

Схема Априорного алгоритма представлена на рис. 1.

В результате работы данного алгоритма выявляются правила, существующие в исходных данных, а также их значения достоверности и подъема.

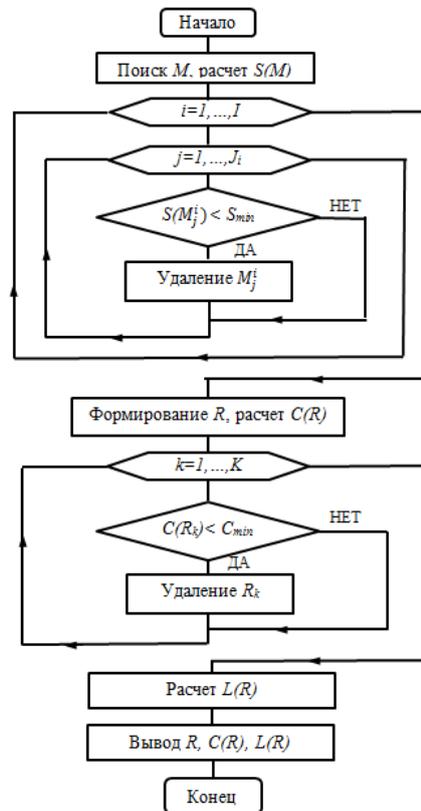


Рис. 1. Схема Априорного алгоритма

2. Модификация Априорного алгоритма ИАД. В Априорном алгоритме значения достоверности и важности определяются на основании частоты встречаемости наборов элементов и правил, в результате чего формируются правила, многие из которых могут быть заранее известными или мало информативными для пользователя ИАД. Однако задача интеллектуального анализа состоит в выявлении, в первую очередь, скрытых закономерностей. Кроме того, при большом объ-

еме полученных правил усложняется процесс их анализа пользователем и выявления среди них наиболее интересных. Для решения этой проблемы в данной работе предлагается ввести дополнительный показатель для найденных правил – их **информативность** для пользователя. Показатель информативности характеризует правила с точки зрения их актуальности для пользователя. Его использование позволит выявлять и выводить, в первую очередь, те правила, которые действительно могут быть скрытыми и, как следствие, актуальными для пользователя.

На успеваемость обучаемых могут влиять различные факторы как относящиеся, так и не относящиеся к учебному процессу. Можно выделить большое число таких факторов, рассматривая учебный процесс с различных точек зрения. В данной работе предлагается использовать следующие основные факторы, оказывающие влияние на результаты учебного процесса:

- 1) со стороны обучаемых:
 - ◆ уровень начальных знаний;
 - ◆ публикационная активность;
- 2) со стороны ППС:
 - ◆ преподаватель, ведущий дисциплину;
 - ◆ использование демонстрационных материалов;
 - ◆ технология контроля;
- 3) со стороны учебно-методического комплекса:
 - ◆ используемые лекционные материалы;
 - ◆ используемые практические материалы;
 - ◆ материалы для оценки знаний;
- 4) со стороны материально-технической базы:
 - ◆ доступность учебно-методических материалов;
 - ◆ обеспеченность оборудованием для обучения;
- 5) со стороны изучаемой дисциплины:
 - ◆ наименование дисциплины;
 - ◆ число часов изучения дисциплины.

Природа этих данных весьма разнообразна, как разнообразен и характер их влияния на успеваемость. Можно предположить, что нет прямой зависимости между некоторыми указанными факторами и уровнем успеваемости, а имеется скорее обратная зависимость. Так, вполне логичным видится утверждение, что студент с лучшей успеваемостью с большей вероятностью будет проявлять публикационную активность, чем его менее заинтересованный в обучении сокурсник. Однако привлечение к научной деятельности, в том числе и менее успевающих студентов, может вызвать у некоторых из них больший интерес к обучению, чем наблюдался ранее. Таким образом, может иметь место прямая зависимость между указанным фактором и показателем успеваемости. С целью дальнейшего расчета значений показателя информативности выявленных правил, все вышеперечисленные факторы разбиты на пять информативных групп в соответствии с рассматриваемой точкой зрения.

Каждой группе присвоим **вес**, характеризующий интерес рассмотрения этой группы элементов пользователем. Обозначим: V – вес группы, V_i – вес i -й группы. Так как интересы пользователя могут меняться в зависимости от решаемых им задач, то величина V является переменной и должна задаваться пользователем в процессе ИАД.

Показатель информативности правил обозначим $Inf(A \rightarrow B)$ и будем рассчитывать его значение на основании информативности соответствующего этому правилу набора элементов $Inf(A)$. Показатель информативности набора элементов будем рассчитывать по формуле:

$$Inf(A) = (\sum_{j=1}^{Ja} Inf(A_j)) / Ja,$$

где $Inf(A_j)$ – информативность j -го элемента набора A ; Ja – число элементов в наборе A .

Информативность отдельного элемента набора определяется в соответствии с весом группы, в состав которой входит этот элемент, по формуле:

$$Inf(A_j) = V_i.$$

Тогда информативность правила определяется следующим образом:

$$Inf(A \rightarrow B) = Inf(A).$$

Полученные значения показателя информативности учитываются при выводе правил пользователю так, что в первую очередь выводятся наиболее информативные правила. Это дает возможность пользователю получить, в первую очередь, именно ту информацию, которая является для него наиболее актуальной.

Как говорилось выше, показатель подъема характеризует наличие обратной зависимости между исходом и посылкой правила вида $A \rightarrow B$. Так как такого рода зависимость в случае образовательных данных в общем случае не имеет смысла, предлагается не рассчитывать величину показателя подъема для правил.

В соответствии с введенными показателями веса групп элементов и информативности правил, а также исключением использования показателя подъема, модифицированный Априорный алгоритм примет вид, представленный на рис. 2.

Предложенная Модификация априорного алгоритма упрощает анализ данных учебного процесса, который характеризуется разнородностью исходных данных, представляющих различный интерес для пользователя. Учет информативности правил позволит выводить пользователю именно те из них, которые являются актуальными, и минимизировать время на просмотр и анализ результатов интеллектуального анализа данных.

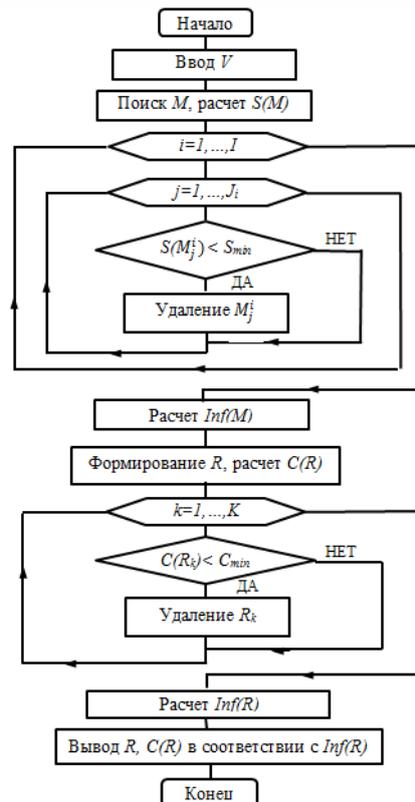


Рис. 2. Модифицированный Априорный алгоритм

3. Использование модифицированного Априорного алгоритма в информационной системе поддержки образовательной деятельности (ИС ПОД). Модификация Априорного алгоритма для анализа данных учебного процесса разработана в рамках ИС ПОД, в состав которой входят: Подсистема оперативного контроля текущей успеваемости студентов [16–17]; Подсистема анализа качества банка тестовых материалов [18]; Подсистема анализа информационно-методического обеспечения учебного процесса [19]; Подсистема ИАД учебного процесса [20]. В рамках последней из перечисленных подсистем и будет использоваться предложенный в данной работе алгоритм.

Архитектура Подсистемы ИАД учебного процесса представлена на рис. 3.

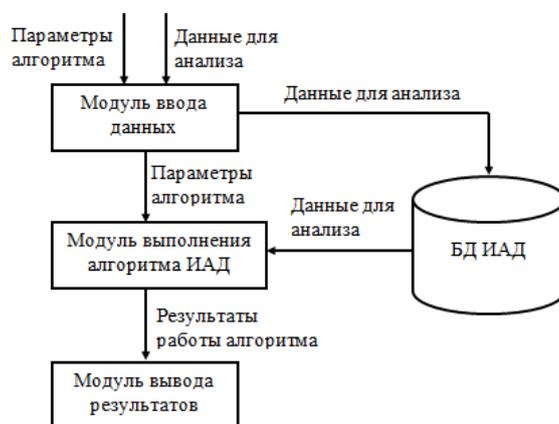


Рис. 3. Архитектура Подсистемы ИАД учебного процесса

Внедрение предложенного алгоритма в подсистему позволит реализовывать ИАД учебного процесса с учетом специфики образовательных данных, что в свою очередь повысит актуальность и информативность результатов анализа.

4. Экспериментальные исследования. Исследование эффективности предложенной модификации Априорного алгоритма для анализа данных учебного процесса в сравнении с исходным алгоритмом [10] производилось с точки зрения повышения информативности результатов ИАД. Для этого был проведен ряд экспериментов, включающий применение исходного Априорного алгоритма и его модификации к одним и тем же исходным данным и сравнение результатов их работы. Рассмотрим пример такого эксперимента.

Пусть имеются исходные данные, представляющие собой таблицу, состоящую из 13 столбцов: 12 – в соответствии с выделенными выше данными для анализа и 1 – успеваемость – анализируемый столбец. Таблица исходных данных содержит 100 строк.

Пользователь присвоил следующие веса группам элементов:

$V_1=0,6$ (факторы со стороны обучаемого),

$V_2=0,4$ (факторы со стороны преподавателя),

$V_3=1$ (факторы со стороны учебно-методического комплекса),

$V_4=0,1$ (факторы со стороны материально-технической базы),

$V_5=0,1$ (факторы со стороны изучаемой дисциплины).

Таким образом, наибольший интерес для него представляет зависимость успеваемости от используемых лекционных, практических и оценочных материалов. Меньший интерес представляют факторы со стороны обучаемого и преподавателя. А факторы со стороны материально-технической базы и изучаемой дисциплины практически не интересуют.

Для введенных данных были выявлены взаимосвязи с учетом весов групп элементов и показателя информативности (в соответствии с предложенной модификацией Априорного алгоритма) и без них (в соответствии с исходным Априорным алгоритмом). Пример результатов работы предложенной модификации Априорного алгоритма представлен на рис. 4.

Правило	Достоверность	Информативность
(НачЗн=4)+(ЛекМ=Методичка4)+(Преп=Преподаватель2) -> (Усп=5)	0,6	0,67
(НачЗн=4)+(ЛекМ=Методичка1) -> (Усп=3)	0,54	0,8
(ЛекМ=Методичка1)+(Преп=Преподаватель5) -> (Усп=4)	0,5	0,7
(ПракМ=Пабник2)+(ЛекМ=Методичка1)+(Преп=Преподаватель4) -> (Усп=4)	0,62	0,8
(ОцМ=БТЗ1)+(ЛекМ=Методичка2)+(ПракМ=Пабник1) -> (Усп=4)	0,85	1
(НачЗн=3)+(ЛекМ=Методичка2)+(Преп=Преподаватель2) -> (Усп=4)	0,55	0,67
(ОцМ=БТЗ2)+(ЛекМ=Методичка1)+(ПубАк=0) -> (Усп=3)	0,45	0,87
(НачЗн=5)+(ЛекМ=Методичка2)+(Преп=Преподаватель5) -> (Усп=5)	0,75	0,67
(НачЗн=4)+(ОцМ=БТЗ2) -> (Усп=3)	0,88	0,7
(ПракМ=Пабник2)+(ЛекМ=Методичка1)+(Преп=Преподаватель5) -> (Усп=5)	0,61	0,8
(ПракМ=Пабник1)+(ЛекМ=Методичка1) -> (Усп=4)	0,45	1

Рис. 4. Пример результатов работы предложенной модификации Априорного алгоритма

Результатом работы обоих алгоритмов являются наборы выявленных закономерностей, однако состав правил в этих наборах отличается. Результаты работы алгоритмов оценивались по следующим критериям: Время работы алгоритма, Функциональность, Количество выявленных правил, Количество информативных правил, Количество неинформативных правил, Актуальность результатов, Оценка удобства поиска интересующих правил. Актуальность результатов рассчитывается как отношение информативных правил к общему числу выявленных правил. Результаты сведены в табл. 1. Был проведен ряд подобных экспериментов, результаты которых оказались аналогичны.

Таблица 1

Показатель	Априорный алгоритм	Предложенная модификация Априорного алгоритма
Время работы алгоритма	< 1 мин.	< 1 мин.
Функциональность	Аналогичная	
Количество выявленных правил	75	43
Количество информативных правил	42	42
Количество неинформативных правил	21	0
Актуальность результатов	0,56	0,98
Оценка удобства поиска интересующих правил	Затрудняется при росте количества правил	Все правила являются информативными

Из табл. 1 видно, что функциональность и производительность предложенной модификации Априорного алгоритма не уступает его оригиналу, однако актуальность результатов его работы выше, что подтверждает его эффективность и целесообразность использования.

Заключение. В работе представлен модифицированный Априорный алгоритм ИАД учебного процесса, учитывающий информационные потребности пользователя. Для этого в алгоритм введен показатель информативности правил, который рассчитывается на основании оценки пользователем степени актуальности рассмотрения тех или иных начальных данных, сгруппированных по степени близости. Введение данного показателя позволяет находить и выводить пользователю в первую очередь те правила, которые представляют для него наибольший интерес. Это позволит не только сократить временные расходы на анализ результатов ИАД, но и исключает случай, когда пользователь может не заметить какое-либо актуальное правило в большом объеме данных.

Внедрение данного алгоритма в ИС ПОД позволит повысить эффективность данной системы за счет добавления возможности реализации в ее рамках ИАД учебного процесса с учетом информационных потребностей пользователя.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Курейчик В.М., Полковникова Н.А. Интеллектуальный анализ многомерных баз данных в экспертной системе // Труды Конгресса по интеллектуальным системам и информационным технологиям «IS&IT'14» (2-9 сентября 2014 г.). – М.: Физматлит, 2014. – Т. 1. – С. 278-286.
2. Курейчик В.М., Полковникова Н.А. Об интеллектуальном анализе баз данных для экспертной системы // Информатика, вычислительная техника и инженерное образование. – 2013. – № 2 (13). – С. 39-50.
3. Witten Ian H., Frank Eibe and Hall Mark A. Data Mining: Practical Machine Learning Tools and Techniques. – 3rd Edition. – Morgan Kaufmann, 2011. – 664 p.
4. Чубукова И.А. Data Mining: учебное пособие. – М.: Интернет-университет информационных технологий: БИНОМ: Лаборатория знаний, 2006. – 382 с.
5. Ситник В.Ф., Краснюк М.Т. Интеллектуальный анализ данных (дейтамайнинг): Навч. посібник. – К.: КНЕУ, 2007. – 376 с.
6. Yang, Y., Guan, H., You, J. CLOPE: A fast and Effective Clustering Algorithm for Transactional Data In Proc. of SIGKDD'02, July 23-26, 2002, Edmonton, Alberta, Canada.
7. Барсегян А.А., Куприянов М.С., Степаненко В.В., Холод И.И. Методы и модели анализа данных: OLAP и Data Mining. – СПб.: БХВ-Петербург, 2004. – 336 с.
8. MacLennan J., Tang Zhi, Crivat B. Data mining with Microsoft SQL Server 2008. – Wiley, 2008. – 672 p.
9. Бова В.В., Заммоев А.У., Дуккардт А.Н. Эволюционная модель интеллектуального анализа разнородных знаний // Известия Кабардино-Балкарского научного центра РАН. – 2013. – № 4 (54). – С. 7-13.
10. Agrawal R., and Srikant R. Fast algorithms for mining association rule // in Proceedings of the 20th International Conference on Very Large Databases. – 1994. – P. 487-499.
11. Рекомендательные системы. Анализ потребительских предпочтений. – URL: http://edu.kpfu.ru/pluginfile.php/75507/mod_resource/content/2/Рекомендательные%20системы_анализ_потребительской_корзины_ноябрь_6_2015.pdf.
12. Технический справочник по алгоритму взаимосвязей (Майкрософт). – URL: [https://msdn.microsoft.com/ru-ru/library/cc280428\(v=sql.120\).aspx](https://msdn.microsoft.com/ru-ru/library/cc280428(v=sql.120).aspx).
13. Шмаков А. Комментарии к статье Ч. Бергера «Data Mining от Oracle: настоящее и будущее». – 2008. – URL: http://citforum.ru/database/oracle/dmsol_comment/.
14. Руководство пользователя SAP Predictive Analysis. – URL: http://help.sap.com/businessobject/product_guides/SAPpa10/ru/pa1_0_8_user_ru.pdf.
15. Авдеев А.Д. Анализ ассоциаций данных потребительской корзины // Теория и практика системного анализа: Труды III Всероссийской научной конференции молодых ученых с международным участием. – Т. II. – Рыбинск: РГАТУ имени П. А. Соловьева, 2014. – С. 5-10.
16. Нужнов Е.В., Казмина И.И. Подсистема оперативного контроля учебного процесса выпускающей кафедры // Известия ЮФУ. Технические науки. – 2013. – № 7 (144). – С. 253-258.
17. Казмина И.И. Нужнов Е.В. Развитие подсистемы оперативного контроля текущей успеваемости студентов // Открытое образование. – 2014. – № 3 (104). – С. 36-41.

18. Казмина И.И., Нужнов Е.В. Трехуровневый анализ качества тестовых материалов // Информатика, вычислительная техника и инженерное образование: Электронный журнал. – 2015. – № 3 (23). – URL: [http://digital-mag.tti.sfedu.ru/lib/13/8-2013-3\(13\).pdf](http://digital-mag.tti.sfedu.ru/lib/13/8-2013-3(13).pdf).
19. Казмина И.И. Организация контроля информационно-методического обеспечения учебного процесса // Информатика, вычислительная техника и инженерное образование: Электронный журнал. – 2013. – № 3 (13). – URL: [http://digital-mag.tti.sfedu.ru/lib/13/8-2013-3\(13\).pdf](http://digital-mag.tti.sfedu.ru/lib/13/8-2013-3(13).pdf).
20. Казмина И.И., Нужнов Е.В. Интеллектуальный анализ данных в образовательной деятельности на основе алгоритма взаимосвязей // Информатика, вычислительная техника и инженерное образование: Электронный журнал. – 2015. – № 1 (21). – URL: [http://digital-mag.tti.sfedu.ru/lib/13/8-2013-3\(13\).pdf](http://digital-mag.tti.sfedu.ru/lib/13/8-2013-3(13).pdf).

REFERENCES

1. Kureychik V.M., Polkovnikova H.A. Intellektual'nyy analiz mnogomernykh baz dannykh v ekspertnoy sisteme [Mining multidimensional databases in the expert system], *Trudy Kongressa po intellektual'nym sistemam i informatsionnym tekhnologiyam «IS&IT'14» (2-9 sentyabrya 2014 g.)* [Proceedings of Congress on intelligent systems and information technologies "IS&IT'14" (2-9 September 2014)]. Moscow: Fizmatlit, 2014, Vol. 1, pp. 278-286.
2. Kureychik V.M., Polkovnikova H.A. Ob intellektual'nom analize baz dannykh dlya ekspertnoy sistemy [About the mining of databases for expert systems], *Informatika, vychislitel'naya tekhnika i inzhenernoe obrazovanie* [Computer Science, Computer Engineering and Engineering Education], 2013, No. 2 (13), pp. 39-50.
3. Witten Ian H., Frank Eibe and Hall Mark A. Data Mining: Practical Machine Learning Tools and Techniques. 3rd Edition. Morgan Kaufmann, 2011, 664 p.
4. Chubukova I.A. Data Mining: uchebnoe posobie [Data Mining: a tutorial]. Moscow: Internet-universitet informatsionnykh tekhnologiy: BINOM: Laboratoriya znaniy, 2006, 382 p.
5. Sitnik V.F., Krasnyuk M.T. Intellektual'nyy analiz danikh (deytamayning): Navch. Posibnik [Data mining (datamining): Proc. Allowance]. Kiev: KNEU, 2007, 376 p.
6. Yang, Y., Guan, H., You, J. CLOPE: A fast and Effective Clustering Algorithm for Transactional Data In Proc. of SIGKDD'02, July 23-26, 2002, Edmonton, Alberta, Canada.
7. Barsegyan A.A., Kupriyanov M.S., Stepanenko V.V., Kholod I.I. Metody i modeli analiza dannykh: OLAP i Data Mining [Methods and models of data analysis: OLAP and Data Mining]. St. Petersburg: BKhV-Peterburg, 2004, 336 p.
8. MacLennan J., Tang Zhi, Crivat B. Data mining with Microsoft SQL Server 2008. Wiley, 2008, 672 p.
9. Bova V.V., Zammoev A.U., Dukkardt A.N. Evolyutsionnaya model' intellektual'nogo analiza raznorodnykh znaniy [An evolutionary model for intelligent analysis of heterogeneous knowledge], *Izvestiya Kabardino-Balkarskogo nauchnogo tsentra RAN* [Izvestija Kabardino-Balkarskogo Nauchnogo Centra RAN], 2013, No. 4 (54), pp. 7-13.
10. Agrawal R., and Srikant R. Fast algorithms for mining association rule, in *Proceedings of the 20th International Conference on Very Large Databases*, 1994, pp. 487-499.
11. Rekomendatel'nye sistemy. Analiz potrebitel'skikh predpochteniy [Recommender systems. Analysis of consumer preferences]. Available at: http://edu.kpfu.ru/pluginfile.php/75507/mod_resource/content/2/Рекомендательные%20системы_анализ_потребительской_корзины_ноябрь_6_2015.pdf.
12. Tekhnicheskii spravochnik po algoritmu vzaimosvyazey (Maykrosoft) [Technical reference Association algorithm (Microsoft)]. Available at: [https://msdn.microsoft.com/ru-ru/library/cc280428\(v=sql.120\).aspx](https://msdn.microsoft.com/ru-ru/library/cc280428(v=sql.120).aspx).
13. Shmakov A. Kommentarii k stat'e Ch. Bergera «Data Mining ot Oracle: nastoyashchee i budushchee» [Comments on the article H Berger "Data Mining from Oracle: present and future"], 2008. Available at: http://citforum.ru/database/oracle/dmsol_comment/.
14. Rukovodstvo pol'zovatelya SAP Predictive Analysis [User guide SAP Predictive Analysis]. Available at: http://help.sap.com/businessobject/product_guides/SAPpa10/ru/pa1_0_8_user_ru.pdf.
15. Avdeev A.D. Analiz assotsiatsiy dannykh potrebitel'skoy korziny [The Association analysis data of the consumer basket], *Teoriya i praktika sistemnogo analiza: Trudy III Vserossiyskoy nauchnoy konferentsii molodykh uchenykh s mezhdunarodnym uchastiem* [Theory and practice of systems analysis: proceedings of the III all-Russian scientific conference of young scientists with international participation]. Vol. II. Rybinsk: RGATU imeni P. A. Solov'eva, 2014, pp. 5-10.

16. *Nuzhnov E.V., Kazmina I.I.* Podsystema operativnogo kontrolya uchebnogo protsessa vypuskayushchey kafedry [The training process on-line monitoring subsystem for profiling department], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2013, No. 7 (144), pp. 253-258.
17. *Kazmina I.I. Nuzhnov E.V.* Razvitie podsystemy operativnogo kontrolya tekushchey uspevaemosti studentov [Development of the subsystem of operational control of current progress of students], *Otkrytoe obrazovanie* [Open Education], 2014, No. 3 (104), pp. 36-41.
18. *Kazmina I.I. Nuzhnov E.V.* Trekhurovnevyy analiz kachestva testovykh materialov [A three-level analysis of the quality of testovi materials], *Informatika, vychislitel'naya tekhnika i inzhenernoe obrazovanie: Elektronnyy zhurnal* [Science, computing and engineering education: the Electronic journal], 2015, No. 3 (23). Available at: [http://digital-mag.tti.sfedu.ru/lib/13/8-2013-3\(13\).pdf](http://digital-mag.tti.sfedu.ru/lib/13/8-2013-3(13).pdf).
19. *Kazmina I.I.* Organizatsiya kontrolya informatsionno-metodicheskogo obespecheniya uchebnogo protsessa [The organization of the control information and methodological support of educational process], *Informatika, vychislitel'naya tekhnika i inzhenernoe obrazovanie: Elektronnyy zhurnal* [Information, computing and engineering education: the Electronic journal], 2013, No. 3 (13). Available at: [http://digital-mag.tti.sfedu.ru/lib/13/8-2013-3\(13\).pdf](http://digital-mag.tti.sfedu.ru/lib/13/8-2013-3(13).pdf).
20. *Kazmina I.I. Nuzhnov E.V.* Intellektual'nyy analiz dannykh v obrazovatel'noy deyatel'nosti na osnove algoritma vzaimosvyazey [Data mining in educational activities based on the Association algorithm], *Informatika, vychislitel'naya tekhnika i inzhenernoe obrazovanie: Elektronnyy zhurnal* [Computer science, computer engineering and engineering education: the Electronic journal], 2015, No. 1 (21). Available at: [http://digital-mag.tti.sfedu.ru/lib/13/8-2013-3\(13\).pdf](http://digital-mag.tti.sfedu.ru/lib/13/8-2013-3(13).pdf).

Статью рекомендовала к опубликованию д.т.н., профессор Л.С. Лисицына.

Казмина Ирина Игоревна – Южный федеральный университет; e-mail: kazmina.i.i@mail.ru; 347928, г. Таганрог, пер. Некрасовский, 44; тел.: 88634371651; кафедра систем автоматизированного проектирования; ассистент.

Нужнов Евгений Владимирович – e-mail: nev@tgn.sfedu.ru; тел.: 88634681885; кафедра систем автоматизированного проектирования; к.т.н.; профессор.

Kazmina Irina Igorevna – Southern Federal University; e-mail: kazmina.i.i@mail.ru; 44, Nekrasovskiy, Taganrog, 347928, Russia; phone: +78634371651; the department of computer aided design; assistant.

Nuzhnov Evgenii Vladimirovich – e-mail: nev@tgn.sfedu.ru; phone: +78634681885; the department of computer aided design; associate professor, professor.

УДК 621.396

DOI 10.18522/2311-3103-2016-7-3952

А.В. Гетманчук

**ВЫСОКОПРОИЗВОДИТЕЛЬНЫЕ МЕТОД И АЛГОРИТМЫ
АВТОМАТИЧЕСКОЙ КЛАССИФИКАЦИИ ОБЪЕКТОВ В УСЛОВИЯХ
ПАРАМЕТРИЧЕСКОЙ НЕОПРЕДЕЛЕННОСТИ И ПЕРЕСЕЧЕНИЯ
КЛАССОВ НА ОСНОВЕ МЕТОДОЛОГИИ С СИСТЕМНОЙ
МАКСИМИЗАЦИЕЙ ЭНТРОПИИ**

Рассматривается задача классификации объектов в условиях параметрической неопределенности и пересечения классов. Спецификой данной задачи является отсутствие достоверной информации о параметрах априорно известных классов, представленных в каталоге эталонных значений в виде доверительных интервалов признаков. Такой вид неопределенности накладывает существенные ограничения на круг методов и алгоритмов, которые могут быть использованы для решения поставленной задачи. В работе рассмотрен метод Г.В. Шелейховского классификации объектов, учитывающий предъявляемые ограничения и обладающий высокой степенью достоверности классификации. Приведены основные недостатки данного метода, затрудняющие его применение в реальных условиях. На основе метода