

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

Г.В. Шелейховского разработан комбинированный метод классификации, обладающий повышенной производительностью и лишенный известных недостатков основополагающего метода. Комбинированный метод предполагает усовершенствование базового метода Г.В. Шелейховского путем анализа взаимного расположения в параметрическом пространстве объектов классификации, представленных в виде точек и априорно известных классов из каталога эталонных значений, представляющих собой области в параметрическом пространстве. В работе представлены следующие алгоритмы, реализующие разработанный метод: алгоритм предварительного анализа входных данных, алгоритм сокращения классификационной матрицы, алгоритм списочного представления классификационной матрицы, алгоритм анализа результатов классификации и формирования информационных данных новых объектов. Разработанный метод позволяет заметно сократить вычислительную трудоемкость базового метода, а также избежать его от ряда существенных недостатков. Так, например, решается известная проблема сходимости метода Г.В. Шелейховского. В статье обосновано существование особых объектов, которые влияют на процесс сходимости метода. Предложенный способ решения проблемы сходимости базового метода классификации отличается от известных тем, что не требует увеличения затрат вычислительных и временных ресурсов. В работе описаны проведенные экспериментальные исследования и представлены полученные результаты. Данные результаты позволяют сделать вывод о снижении общей вычислительной трудоемкости при классификации объектов с использованием комбинированного метода по отношению к методу классификации Г.В. Шелейховского в среднем в два раза.

Классификация объектов; неопределенность; реальное время; энтропия; нормирование; доверительный интервал; классификационная матрица.

A.V. Getmanchuk

HIGH-EFFICIENCY METHOD AND ALGORITHMS OF AUTOMATIC CLASSIFICATION OF OBJECTS IN THE CONDITIONS OF PARAMETRICAL UNCERTAINTY AND CLASS INTERSECTION ON THE BASIS OF METHODOLOGY WITH SYSTEM MAXIMIZATION OF ENTROPY

In the report the problem of objects classification in the conditions of parametrical uncertainty and class intersection is considered. Specificity of the problem is absence of reliable information about parameters of a priori known classes presented in the catalogue of reference values in the form of confidence intervals of signs. Such kind of uncertainty imposes essential restrictions on a range of methods and algorithms which can be used for the solution of the problem. In the report G.V. Sheleykhovskiy's method of objects classification, which takes into account required restrictions and possessing high degree of reliability of classification is considered. The basic imperfections of the method complicating its application in real conditions are shown. On the basis of G.V. Sheleykhovskiy's method the combined method of the classification possessing improved productivity and devoid of known imperfections of the basic method is developed. The combined method assumes improvement of the base G.V. Sheleykhovskiy's method by the relative positioning analysis in parametrical space of the classification objects presented in the form of points and a priori known classes from the catalogue of reference values, representing areas in the parametrical space. In the document the following algorithms realising the developed method are presented: algorithm of the entrance data preliminary analysis, algorithm of the classification matrix reduction, algorithm of list representation of the classification matrix, algorithm of classification results analysis and formation of information data of new objects. The developed method allows to reduce considerably computing labouriousness of a base method, and also to relieve it of a number of essential defects. So, for example, the known problem of G. V. Sheleykhovskiy's method convergence is solved. In article existence of specific objects which influence process of convergence of this method is proved. The offered way of a solution of a convergence problem of a base method of classification differs from known ways that it does not demand increase in expenses of computing and time resources. In work the performed experimental researches are described and the received results are presented. The present results allow to draw a conclusion on decrease in the general of computing labouriousness at classification of objects with use of the combined method in relation to the method of classification of G.V. Sheleykhovskiy on the average twice.

Objects classification; uncertainty; real time; entropy; normalization; confidence interval; a classification matrix.

Введение. Современные тенденции развития науки и техники связаны, в частности, с необходимостью эффективного решения задач классификации объектов в условиях различного рода ограничений в представлении исходных данных, требований к алгоритмической реализации функциональных возможностей и вычислительным средствам при практической реализации. Систему распознавания образов можно представить в упрощенном виде, приведенном на рис. 1.

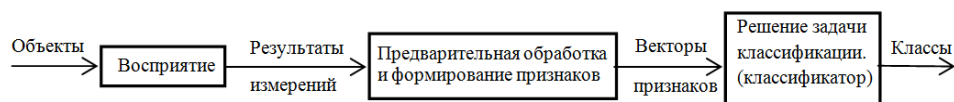


Рис. 1. Структурная схема системы распознавания образов.

Цель работы системы распознавания образов заключается в том, чтобы на основе собранной информации определить класс объектов с характеристиками, аналогичными измеренным у распознаваемых объектов [1]. На рис. 1 блоки «Восприятие» и «Предварительная обработка и формирование признаков» зависят от специфики области применения системы распознавания образов конкретной реализации измерительной аппаратуры, в то время как классификатор оперирует лишь признаками объектов. Часто оказывается полезным рассматривать признаки образов в качестве точек n -мерного евклидова пространства [1]. В рамках данной работы рассматриваются вопросы, охватывающие алгоритмическую организацию решения задачи классификации объектов по их признакам.

1. Постановка задачи. Рассматривается задача классификации объектов в условиях параметрической неопределенности и пересечения классов характеризующаяся следующими особенностями:

1) на вход классификатора в реальном времени с определенной периодичностью поступает множество M признаков одновременно наблюдаемых объектов w ; объекты характеризуются значениями признаков x_i , $i=1, \dots, N$; количество разновидностей объектов измеряется десятками-сотнями;

2) эталонные значения априорно известных классов характеризуются параметрической неопределенностью, которая обусловлена отсутствием достоверной информации о значениях параметров априорно известных классов объектов, неполным перечнем всех возможных классов, а также ограниченной точностью измерения параметров объектов; поэтому в каталоге эталонных значений априорно известных классов для каждого параметра любого класса вводится некоторый доверительный интервал;

3) попадание параметра объекта классификации в доверительный интервал класса говорит о существовании возможности отнесения анализируемого объекта к данному классу с некоторой вероятностью. Причем эту вероятность невозможно оценить исходя из места попадания параметра в интервал относительно границ или центра данного доверительного интервала, так как доверительный интервал является следствием отсутствия достоверной информации о значениях параметров априорно известных классов объектов;

4) в связи с тем, что классы в каталоге эталонных значений представлены доверительными интервалами, существует возможность их пересечения, а следовательно, не исключена ситуация отнесения объекта классификации к более, чем одному классу;

5) используемая методология решения задачи классификации должна характеризоваться возможностью глобальной (единой по всей методологии) критериальной оптимизации в получаемых оценках вероятностей принадлежности объектов к соответствующим классам;

б) каталог эталонных значений классов содержит не полный перечень всех возможных классов, в связи с чем необходимо в непрерывном процессе выявлять закономерности возникновения неизвестных (новых) объектов с целью дополнения указанного каталога (например, в автоматизированном режиме работы системы с привлечением оператора-эксперта).

Необходимость решения задачи классификации, соответствующей указанным требованиям, возникает, например, при классификации радиотехнических объектов, зондировании поверхности земли, при диагностике в биологии и медицине, прогнозировании в геологии. Обычно задачи подобного рода возлагаются на АСУ различных комплексов и относятся к рангу задач оценки состояния объекта управления [2, 3]. Из этого следует, что вопрос минимизации вычислительной нагрузки методов, как правило, решающих данные задачи в реальном времени, является актуальным и очень важным. Целью работы является разработка метода и алгоритмических средств высокопроизводительной классификации объектов в условиях неопределенности, характеризующейся отсутствием достоверной информации о параметрах априорно известных классов, основанного на принципе системной максимизации энтропии.

2. Обзор известных методов классификации. Известны методы классификации объектов, в частности с пересечением классов, среди которых можно выделить статистические, на основе нечеткой логики, нейросетевые, детерминистские [1–7]. С учетом особенностей сформулированной выше задачи классификации применение статистического подхода представляется проблематичным в связи с представлением эталонов каталога без использования статистических свойств классов [1].

Детерминистский подход включает в себя различные методы [4]. Среди детерминистских методов наиболее развитыми являются метрические методы диагностирования, основанные на понятии расстояния между классами [5]. Здесь для построения решающих правил нужна обучающая выборка. Обучающая выборка – это множество объектов, заданных значениями признаков и принадлежность которых к тому или иному классу достоверно известна "учителю" и сообщается учителем "обучаемой" системе. По обучающей выборке система строит решающие правила. Важной особенностью (недостатком) детерминистских методов решения задач распознавания является условие обеспечения достаточно полного представления генеральной совокупности обучающей и экзаменационной выборками (гипотетическое множество всех возможных объектов каждого образа), что не представляется возможным в условиях отсутствия достоверной информации о параметрах априорно известных классов [8–10].

В алгоритмах, использующих нечеткую логику, нечеткое множество строго определяется с помощью функции принадлежности [6, 7]. Однако в соответствии со спецификой постановки задачи, рассматриваемой в рамках данной работы, исключается возможность создания данной функции принадлежности. Это связано с тем, что попадание параметров объекта классификации в доверительные интервалы класса из каталога эталонных значений говорит лишь о существовании вероятности отнесения анализируемого объекта к данному классу. Данная вероятность неизвестна и никак не связана с геометрическим местом попадания параметра в доверительный интервал относительно его границ. Кроме того, известные методы базируются на трудоемких алгоритмах полного перебора.

При использовании нейросетевого подхода классификации необходимо учитывать тот факт, что «не существует гарантии, что нейронная сеть может быть обучена за конечное время» [12]. Создание обучающей выборки проблематично в связи с наличием параметрической неопределенности классов, выраженной в виде

доверительных интервалов каталога эталонных значений. Проблематично также переобучение сети при интенсивном периодическом поступлении для классификации, в частности, новых объектов, а также коррекции каталога эталонов.

Следует также отметить, что в указанных выше методах отсутствует единый системный критерий оптимизации результатов классификации. Известные методы, кроме того, в условиях интенсивного поступления для одновременной классификации десятков – сотен объектов характеризуются высокой вычислительной трудоемкостью, создающей существенные проблемы для организации работы системы в реальном времени.

В связи с отмеченным представляет интерес выбор такой методологии для решения рассматриваемой задачи классификации объектов, которая позволила бы разрешить описанные проблемы.

В рамках данной работы для решения задачи классификации объектов предлагается использовать нетрадиционный так называемый метод Г.В. Шелейховского. Этот метод еще в 30-х годах (прошлого столетия) был предложен ленинградским архитектором Г.В. Шелейховским для расчета пассажиропотоков [13]. В [14] показано, что для более точного определения степени достоверности классификации следует применять метод классификации Г.В. Шелейховского.

Метод Г.В. Шелейховского предполагает проводить распознавание одновременно наблюдаемых объектов как на основе сравнения замеренных параметров с параметрами априорно известных классов из каталога, так и на основе сравнения набора параметров всех одновременно наблюдаемых объектов между собой. Суть метода заключается в следующем.

Для каждого объекта классификации с измеренными признаками (параметрами) из каталога эталонных значений выбираются классы, которым может соответствовать анализируемый объект. На данном этапе каждому объекту классификации приписывается перечень возможных классов, к которым данный объект может быть отнесен. При этом класс «новый» приписывается как возможный к каждому объекту в обязательном порядке для дополнения результатов классификации до полной группы событий.

На следующем этапе заполняется классификационная матрица $\Delta = |\delta_{kj}|$, $k=1..V$, $j=1..W$. В данной матрице число столбцов «V» равно числу классов из каталога эталонных значений, к которым могут относиться все наблюдаемые объекты классификации. Число строк матрицы «W» равно числу подмножеств однотипных объектов. Сами значения элементов матрицы равны: 1 – если k-объект имеет в своей подборке j-класс из каталога; 0 – если не имеет.

На третьем этапе классификационная матрица дополняется до квадратной, если она не является таковой. Дополнение происходит путем добавления к матрице недостающих строк, либо столбцов, состоящих из единиц. Каждому столбцу присваивается имя «новый», каждой строке – «фиктивная».

На следующем этапе циклично производится поочередное выполнение двух операций:

1. Нормирование по строкам, т.е. деление каждого элемента матрицы на сумму элементов строки, в которой стоит данный элемент:

$$\delta_{k,j}(t+1) = \delta_{k,j}(t) / \sum \delta_{k,j}(t), \quad (1)$$

где $\delta_{k,j}(t)$ – значение элемента матрицы Δ , стоящего в k-й строке и j-м столбце после t-го шага процедуры (до операции транспонирования).

1. Нормирование по столбцам:

$$\delta_{k,j}(t+1) = \delta_{k,j}(t) / \sum \delta_{k,j}(t). \quad (2)$$

Данная процедура последовательного нормирования носит название процедуры Г.В. Шелейховского. Доказано [10], что процедура Г.В. Шелейховского, применяемая к матрице $\Delta = |\delta_{k,j}|$ устремляет каждый элемент матрицы $\delta_{k,j}$ к некоторому значению $r_{k,j}$, причем:

$$\sum_k p_{k,j} \rightarrow 1; \sum_j p_{k,j} \rightarrow 1; -\sum_k \sum_j p_{k,j} * \log_2(p_{k,j}) \rightarrow \max. \quad (3)$$

Итерации повторяются до тех пор, пока суммы элементов каждой строки не будут отличаться от единицы более, чем на 0,001. Процедура Г.В. Шелейховского, примененная к классификационной матрице, преобразует последнюю в матрицу вероятностей принадлежности наблюдаемых объектов к априорно известным, или новым классам. При этом, полученное распределение максимизирует энтропию системы [13]. Согласно принципу Дженса [15–17], распределение вероятностей, при котором энтропия максимальна, является «наименее сомнительным представлением вероятностей при учете всей заданной информации». Таким образом, метод Г.В. Шелейховского для классификации объектов в условиях параметрической неопределенности и пересечения классов обладает высокой степенью достоверности классификации.

Однако данный метод имеет ряд существенных недостатков, затрудняющих его применение в реальных условиях. К ним относятся высокая вычислительная трудоемкость, а также проблема сходимости, возникающая при некоторых наборах входных данных. Эта проблема подробно описана в работе [18]. Известны подходы к решению данной проблемы [18–20], однако они либо не гарантируют положительного результата, либо заметно усложняют исходный метод. Метод Г.В. Шелейховского также не уделяет должного внимания информации о новых объектах, что лишает возможности дополнения каталога эталонных значений актуальной информацией.

Все вышесказанное говорит об актуальности и научном интересе в разработке и исследовании высокопроизводительных метода и алгоритмов автоматической классификации объектов в условиях параметрической неопределенности и пересечения классов. При этом представляет интерес разработка комбинированного метода классификации, основанного на усовершенствовании метода Г.В. Шелейховского (далее базовый метод) и характеризующегося отсутствием или ослаблением недостатков данного метода.

3. Разработка комбинированного метода классификации. Основная идея разрабатываемого комбинированного метода классификации сводится к тому, чтобы ввести в базовый метод понятие параметрического пространства и представить в нем объекты классификации и классы из каталога эталонных значений. Введение параметрического пространства в базовый метод позволяет добиться улучшения технических характеристик алгоритма, путем применения детерминистских методов, а также методов кластерного анализа.

Разработаны основные алгоритмы, реализующие комбинированный метод классификации:

- ◆ алгоритм предварительного анализа входных данных;
- ◆ алгоритм сокращения классификационной матрицы;
- ◆ алгоритм списочного представления классификационной матрицы;
- ◆ алгоритм анализа результатов классификации и формирования информационных данных новых объектов.

Алгоритм предварительного анализа входных данных представлен в виде двух подэтапов следующим образом:

1) предварительный анализ каталога эталонных значений. По аналогии с алгоритмом дробящихся эталонов строятся гиперпараллелепипеды (далее области) в параметрическом пространстве для каталога эталонных значений. Наличие пересечений данных областей говорит о том, что может возникнуть проблема отнесения анализируемого объекта к более, чем одному классу. В противном случае, если нет ни одного пересечения, можно с уверенностью сказать, что все поступаю-

щие на обработку объекты будут однозначно классифицированы, или в случае непопадания ни в одну область отнесены к классу «новый объект». Следовательно, еще до поступления объектов на обработку, можно установить, будет ли целесообразным применение процедуры последовательного нормирования, или алгоритм «выродится» в алгоритм построения эталонов.

2) если первый шаг подготовительного этапа выявил наличие параметрических пересечений в каталоге эталонных значений, возникает необходимость анализа набора входных данных на предмет попадания параметров анализируемых объектов в параметрические пересечения каталога. В случае попадания хотя бы одного объекта в более чем одну область, принимается решение о применении процедуры последовательного нормирования, в противном случае применение данной процедуры является необоснованным.

Формирование классификационных образов объектов классификации по приведенным выше п.1 и п.2 происходит следующим образом:

- ◆ для очередного объекта классификации ищется его первое попадание в область по всем признакам;
- ◆ если обнаружена область, в которую попадает объект – выбираются области, имеющие пересечения с обнаруженной;
- ◆ анализируется попадание объекта классификации в области, выбранные на предыдущем шаге;
- ◆ составляется список областей, в которые попадает объект – классификационный образ объекта.

Таким образом, классификационная матрица строится не методом последовательного перебора возможности отнесения анализируемого объекта к каждому классу из каталога эталонных значений, а по результатам данного шага алгоритма.

Алгоритм сокращения классификационной матрицы основывается на следующих выводах, полученных в результате анализа работы процедуры последовательного нормирования:

1) если объект $k \in \bigcap_i X_i$, где X_i – некоторый набор областей из состава каталога эталонных значений, то на формирование конечного результата по данному объекту будут влиять все объекты m из группы одновременно наблюдаемых объектов, такие что $m \in \bigcup_i X_i$;

2) можно ввести определение: два объекта $k \in Y$ и $m \in Y$, где Y – группа одновременно наблюдаемых объектов, называются взаимозависимыми, если $\Delta k \cap \Delta m \neq \emptyset$, где $\Delta k = \{\delta_i^k\}$ и $\Delta m = \{\delta_i^m\}$ – классификационные образы объектов k и m . В противном случае объекты являются независимыми;

3) влияние друг на друга независимых объектов из состава группы одновременно наблюдаемых, возможно только в случае присутствия в классификационной матрице единичных элементов, введенных в матрицу с целью дополнения ее до квадратной. Это действие является искусственным и необходимо только для правильной работы процедуры последовательного нормирования. В данном случае влияние друг на друга независимых объектов в пределах классификационной матрицы является косвенным и может не учитываться в силу своей незначительности;

4) с целью сокращения затрат на выполнение процедуры последовательного нормирования, имеет смысл исключить из процесса формирования классификационной матрицы объекты, не имеющие взаимозависимых объектов в группе одно-

временно наблюдаемых. Эти объекты не влияют на распределение вероятности других объектов группы. Оставшееся множество объектов следует разделить на подмножества взаимозависимых объектов и для каждого такого подмножества построить классификационную матрицу.

На рис. 2 приведен упрощенный пример возможного взаимного расположения объектов классификации и классов в двумерном параметрическом пространстве.

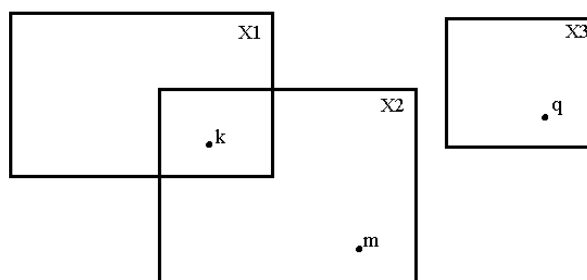


Рис. 2. Пример представления объектов классификации и классов

На рис. 2 представлены классы X_1 , X_2 и X_3 из каталога эталонных значений и объекты классификации k , m и q . Здесь объекты k и m – взаимозависимые, а объект q – независимый

Разработанный алгоритм сокращения позволяет существенно уменьшить размерность классификационной матрицы, а в некоторых случаях разбить ее на несколько подматриц, что заметно снижает нагрузку при выполнении процедуры последовательного нормирования. При этом, во время работы алгоритма формируется список заранее классифицированных объектов. Для объектов из данного списка распределение вероятностей отнесения их к классам, составляющим их классификационные образы, является равномерным.

Суть алгоритма списочного представления классификационной матрицы заключается в исключении из рассмотрения нулевых элементов матрицы, т.е. создании псевдоматрицы, состоящей только из значащих элементов. В роли строки матрицы выступает динамический список, который создается при анализе принадлежности i -го объекта к j -му классу. В случае возможности принадлежности создается очередной элемент списка. Столбцы матрицы также являются списками. Каждый элемент матрицы входит в состав списка – строки и списка – столбца.

Разработанный алгоритм списочного представления классификационной матрицы позволяет при выполнении процедуры классификации перемещаться непосредственно от одного значащего элемента матрицы к другому. Данный алгоритм дает возможность сократить затраты ресурсов ЭВМ, исключая нулевые элементы матрицы не только из обработки, но и не выделяя память для их хранения.

Сбор и обработка информации о новых классах, с целью дальнейшей корректировки каталога, является актуальной и очень важной задачей. Разработанный алгоритм анализа полученных результатов классификации приведен ниже. При первом измерении в уже имеющемся параметрическом пространстве, созданном на подготовительном этапе, для каждого неклассифицированного объекта строится область. Здесь центром области является сам объект, а радиусом – заданное заранее значение. При следующем измерении существование области должно подтвердиться попаданием в нее объекта из состава входных данных. Если существование области подтвердилось определенное количество раз (порог подтверждений), информация об эталоне данной области заносится во временный каталог новых классов, в противном случае область удаляется из параметрического про-

странства. Если информация об эталоне области уже существует в каталоге – она перезаписывается. Для объекта, не попавшего ни в одну область, строится новая область. В результате работы алгоритма оператор АСУ имеет информацию о новых объектах, автоматически обнаруженных во время процесса классификации. Данная информация может быть использована при корректировке каталога эталонных значений. Для персонала АСУ появилась возможность по мере формирования эмпирического знания осуществлять детальную оценку состояния объекта управления и на ее основе корректировать работу АСУ. Таким образом может быть учтено влияние на работу АСУ различных внешних факторов, а также индивидуальные особенности измерительной аппаратуры.

Детальный анализ проблемы сходимости в решении задачи классификации базовым методом показывает, что данная проблема вызвана наличием в группе одновременно наблюдаемых объектов, участвующих в эксперименте, хотя бы одного неоднозначно определенного объекта. Этот объект не имеет объектов, которые могут повлиять на распределение его вероятностей. То есть проблема сходимости возникает в том случае, если существует такой объект, который можно отнести к более, чем одному классу из каталога эталонных значений. Но при этом нет такого объекта в группе одновременно наблюдаемых, который относился бы хотя бы к одному из этих классов. С точки зрения комбинированного метода классификации «проблемный» объект будет характеризоваться попаданием в пересечение нескольких областей и являться при этом единственным объектом, попавшим в данное подмножество пересекающихся областей. Так как «проблемные» объекты не имеют объектов в группе одновременно наблюдаемых, на распределение вероятностей которых они могут повлиять, они могут быть исключены из процесса формирования классификационной матрицы без изменения результатов классификации. Таким образом, предложенное решение проблемы сходимости задачи классификации не только не замедляет работу метода, как другие известные ранее способы [18-20], но наоборот сокращает вычислительную трудоемкость до предельно минимального уровня, соответствующего решению задачи при отсутствии проблемы сходимости. Для исключенного «проблемного» объекта предполагается равномерное распределение вероятностей среди классов, представленных единицами в его классификационном образе. Рассмотренный материал наиболее близко соответствует решению задачи классификации радиотехнических объектов в условиях неопределенности [14].

4. Экспериментальные исследования. В рамках экспериментальных исследований разработана программная модель, представляющая возможности для проведения экспериментов с целью исследования характеристик предлагаемого комбинированного метода классификации и сравнения с характеристиками базового метода. Были проведены следующие экспериментальные исследования: исследование алгоритма предварительного анализа входных данных; анализ применения списочного алгоритма представления классификационной матрицы; оценка влияния сокращения классификационной матрицы на работу алгоритма классификации.

Алгоритм предварительного анализа входных данных на случайных наборах данных показал снижение вычислительной трудоемкости по отношению к базовому методу в среднем в два раза. Результаты проведенного эксперимента представлены на рис. 3.

На рис. 3 прямая линия, соответствующая значению 100 операций сравнения, обозначает количество операций сравнения при использовании базового метода. Каждая точка на графике – результат одного эксперимента с использованием алгоритма предварительного анализа входных данных.

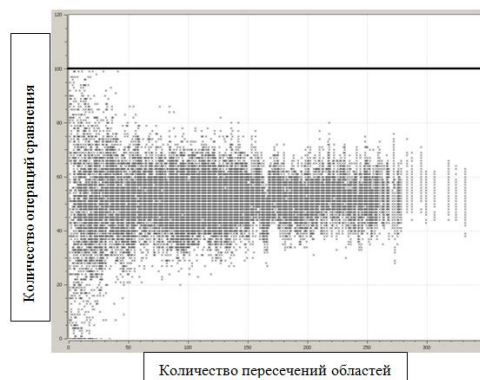


Рис. 3. Результаты работы алгоритма предварительного анализа входных данных

Анализ применения списочного алгоритма представления классификационной матрицы заключался в сравнении нового алгоритма с матричным алгоритмом представления и обработки данных. Здесь на вход алгоритмов подавались автоматически генерируемые данные, соответствующие различным комбинациям заполнения классификационной матрицы единичными элементами. При этом оценивалось количество математических операций при выполнении процедуры последовательного нормирования. Результаты эксперимента представлены на рис. 4 и 5.

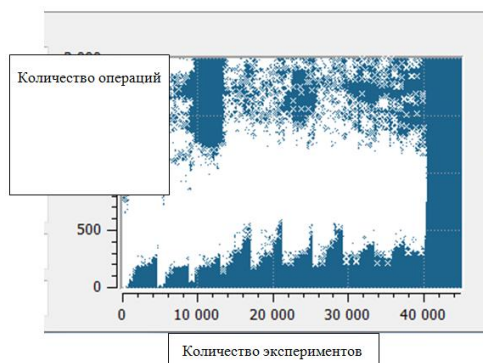


Рис. 4. Отображение результатов работы матричного метода

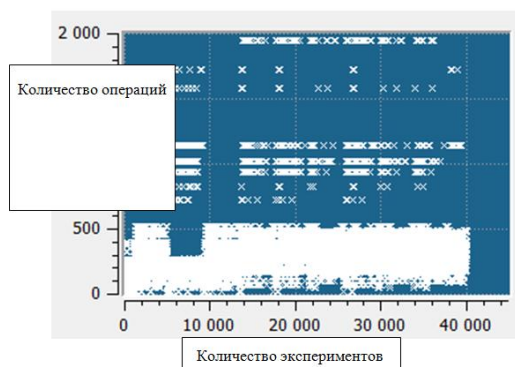


Рис. 5. Отображение результатов работы списочного метода

Проведенный эксперимент показал среднее количество операций, затраченных на выполнение процедуры последовательного нормирования матричным методом, равное 736. Среднее количество операций для списочного метода составило 344.

Для проведения эксперимента по оценке влияния сокращения классификационной матрицы на работу алгоритма классификации были подготовлены несколько наборов входных данных, отличающихся друг от друга количеством пересечений областей в файлах, содержащих параметры каталога эталонных значений. При этом оценивалось количество операций для различных методов классификации. Результаты эксперимента представлены в табл. 1.

Таблица 1

Результаты сокращения классификационной матрицы

	Количество пересечений в файлах областей			
	0	100	200	300
Количество операций, исходная матрица. Матричный способ.	135036	64000	73008	67712
Количество операций, исходная матрица. Списочный способ.	8184	15440	18480	12332
Количество операций, сокращенная матрица. Матричный способ.	-----	18284	32812	28588
Количество операций, сокращенная матрица. Списочный способ.	-----	9200	14699	9920

Проведенный эксперимент является примером возможности сокращения вычислительной трудоемкости алгоритма в среднем в два раза. Результаты эксперимента позволяют сделать следующий вывод: применение алгоритма сокращения классификационной матрицы на подготовительном этапе работы алгоритма существенно сокращает затраты на выполнение процедуры последовательного нормирования. Проведенный эксперимент в результате дал преимущество применения сокращенной матрицы перед исходной по количеству операций сравнения в ~2 раза.

Была выполнена проверка работоспособности алгоритма анализа информации о новых объектах, эксперимент проводился в режиме динамической подачи входных данных.

В рамках экспериментального исследования была проведена проверка работоспособности комбинированного метода в случае возникновения проблемы сходимости задачи классификации. Комбинированный метод показал устойчивость к входным данным, вызывающим проблему сходимости базового метода, в результате чего был получен конечный результат классификации.

В результате проведенных экспериментов были получены статистические оценки, характеризующие снижение вычислительной трудоемкости: количество операций сравнения на подготовительном этапе снижается по статистической оценке в среднем два раза; снижение количества операций сравнения при применении списочного алгоритма по сравнению с матричным по статистическим оценкам снижается в среднем в два раза; при использовании алгоритма сокращения классификационной матрицы снижение вычислительной трудоемкости варьируется в пределах от максимально необходимой при отсутствии возможности сокращения до полного исключения выполнения процедуры последовательного норми-

рования в случае вырождения классификационной матрицы. Результаты проведенных экспериментов позволяют сделать вывод о снижении общей вычислительной трудоемкости при классификации объектов с использованием комбинированного метода в среднем в два раза.

Заключение. Предлагаемый комбинированный метод классификации обладает, как и базовый метод, возможностью обеспечения системной максимизации энтропии. В то же время он характеризуется следующими преимуществами по сравнению с базовым методом:

- 1) на подготовительном этапе работы алгоритма классификации делается вывод о целесообразности применения процедуры последовательного нормирования;
- 2) снижены затраты на выполнение процедуры последовательного нормирования сокращением классификационной матрицы;
- 3) решена проблема сходимости задачи классификации с применением процедуры последовательного нормирования;
- 4) классификационная матрица представлена в виде совокупности динамических списков, что позволяет сократить вычислительную трудоемкость, исключая нулевые элементы матрицы не только из обработки, но и не выделяя память для их хранения;
- 5) расширена функциональность обработки результатов классификации с возможностью дополнения каталога эталонных значений актуальной информацией;
- 6) общая оценка вычислительной трудоемкости разработанного комбинированного метода классификации объектов показывает его преимущество перед известным ранее базовым методом классификации в среднем в два раза.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Гонсалес Р., Ту Дж. Принципы распознавания образов. – М.: Мир, 1978. – 411 с.
2. Симанков В.С., Луценко Е.В. Адаптивное управление сложными системами на основе теории распознавания образов. – Краснодар: Техн. ун-т Кубан. гос. технол. ун-та, 1999. – 318 с.
3. Кириллов Н.П. Признаки класса и определение понятия «технические системы» // Авиа-космическое приборостроение. – 2009. – Вып. 8. – С. 32-38.
4. Журавлев Ю.И. Об алгебраическом подходе к решению задач распознавания и классификации // Проблемы кибернетики. – 1978. – Вып. 33. – С. 5-68.
5. Дорогов А.Ю. Теория и проектирование быстрых перестраиваемых преобразователей и слабосвязанных нейронных сетей. – СПб.: Политехника, 2014. – 344 с.
6. Тэрано Т., Асаи К., Сугэно М. Прикладные нечеткие системы. – М.: Мир, 1993. – 368 с.
7. Кручинин И.И., Логинов Б.М. Моделирование процессов классификации многомерных объектов пересекающихся классов на основе представлений нечетких множеств // Труды 2-й Российской конференции молодых ученых по математическому моделированию. – М.: МГТУ, 2002. – С. 14-23.
8. Волошин Г.Я. Методы распознавания образов конспект лекций. Сайт цифровых учебно-методических материалов ВГУЭС. – Режим доступа: http://abc.vvsu.ru/Books/Metody_r.
9. Журавлёв Ю.И. Распознавание образов и распознавание изображений. Распознавание, классификация, прогноз // Математические методы и их применение. – 1989. – Вып. 2. – С. 5-73.
10. Ванник В.Н., Червоненкис А.Я. Теория распознавания образов. – М.: Наука. 1974. – 416 с.
11. Кручинин И.И., Логинов Б.М. Моделирование процессов классификации многомерных объектов пересекающихся классов на основе гибридных нейросетей // Труды 2-й Российской конференции молодых ученых по математическому моделированию. – М.: МГТУ, 2002. – С. 24-30.
12. Круг П.Г. Нейронные сети и нейрокомпьютеры: учебное пособие по курсу «Микропроцессоры». – М.: Изд-во МЭИ, 2002. – 176 с.
13. Брегман Л.М. Доказательство сходимости метода Шеллейховского для задачи с транспортными ограничениями // Вычислительная математика и математическая физика. – М., 1967. – С. 147-156.

14. *Гришков А.Ф., Кулаков А.А., Шнак В.Ф.* Классификация образов радиотехнических сигналов // Вопросы специальной радиоэлектроники. – 2009. – Вып. 2. – С. 30-37.
15. *Трухаев П.И.* Модели принятия решений в условиях неопределенностей. – М.: Наука, 1981. – 258 с.
16. *Jaynes, E.T.* Information Theory and Statistical Mechanics // Physical Review. – 1957. – No. 106 (4). – P. 620-630. DOI:10.1103/PhysRev.106.620. Bibcode: 1957PhRv..106..620J.
17. *Jaynes E.T.* Information Theory and Statistical Mechanics II // Physical Review. – 1957. – No. 108 (2). – P. 171-190. DOI: 10.1103/PhysRev.108.171. Bibcode: 1957PhRv..108..171J.
18. *Шнак В.Ф., Гетманчук А.В., Радченко С.А., Кулаков А.А.* К вопросу о сходимости алгоритма классификации радиотехнических образов по методу Г.В. Шелейховского // Вопросы специальной радиоэлектроники. – 2011. – Вып. 2. – С. 41-46.
19. *Sinkhorn, Richard.* A relationship between arbitrary positive matrices and doubly stochastic matrices. Ann. Math. Statist, 1964.
20. *Sinkhorn, Richard, & Knopp, Paul.* Concerning nonnegative matrices and doubly stochastic matrices. Pacific J. Math, 1967.

REFERENCES

1. *Gonsales R., Tu Dzh.* Printsipy raspoznavaniya obrazov [Principles of pattern recognition]. Moscow: Mir, 1978, 411 p.
2. *Simankov V.S., Lutsenko E.V.* Adaptivnoe upravlenie slozhnymi sistemami na osnove teorii raspoznavaniya obrazov [Adaptive control of complex systems based on the theory of pattern recognition]. Krasnodar: Tekhn. un-t Kuban. gos. tekhnol. un-ta, 1999, 318 p.
3. *Kirillov N.P.* Priznaki klassa i opredelenie ponyatiya «tekhicheskie sistemy» [Class characteristics and the definition of "technical systems"], *Aviakosmicheskoe priborostroenie* [Aerospace Instrumentation], 2009, Issue 8, pp. 32-38.
4. *Zhuravlev Yu.I.* Ob algebraicheskom podkhode k resheniyu zadach raspoznavaniya i klassifikatsii [About algebraic approach to solving the problems of recognition and classification], *Problemy kibernetiki* [Problems of Cybernetics], 1978, Issue 33, pp. 5-68.
5. *Dorogov A.Yu.* Teoriya i proektirovanie bystrykh perestraivaemykh preobrazovateley i slabosvyazannykh neyronnykh setey [Theory and design fast tunable transmitters and loosely-coupled neural networks]. St. Petersburg: Politehnika, 2014, 344 p.
6. *Terano T., Asai K., Sugeno M.* Prikladnye nechetkie sistemy [Applied fuzzy systems]. Moscow: Mir, 1993, 368 p.
7. *Kruchinin I.I., Loginov B.M.* Modelirovanie protsessov klassifikatsii mnogomernykh ob"ektov peresekayushchikhsya klassov na osnove predstavleniy nechetkikh mnozhestv [Modeling of processes of classification of multidimensional objects overlapping classes based on the concepts of fuzzy sets], *Trudy 2-y Rossiyskoy konferentsii molodykh uchenykh po matematicheskoy modelirovaniyu* [Proceedings of 2-nd Russian conference of young scientists on mathematical modeling]. Moscow: MGTU, 2002, pp. 14-23.
8. *Voloshin G.Ya.* Metody raspoznavaniya obrazov konspekt lektsiy. Sayt tsifrovyykh uchebno-metodicheskikh materialov VGUES [Pattern recognition lecture notes. The website of digital educational and methodological materials VSUES]. Available at: http://abc.vvsu.ru/Books/Metody_r.
9. *Zhuravlev Yu.I.* Raspoznavanie obrazov i raspoznavanie izobrazheniy. Raspoznavanie, klassifikatsiya, prognoz [Pattern recognition and image recognition. Recognition, classification, prediction], *Matematicheskie metody i ikh primeneniye* [Mathematical Methods and their Application], 1989, Issue 2, pp. 5-73.
10. *Vapnik V.N., Chervonenkis A.Ya.* Teoriya raspoznavaniya obrazov [Theory of pattern recognition]. Moscow: Nauka. 1974, 416 p.
11. *Kruchinin I.I., Loginov B.M.* Modelirovanie protsessov klassifikatsii mnogomernykh ob"ektov peresekayushchikhsya klassov na osnove gibridnykh neyrosetey [Modeling of processes of classification of multidimensional objects overlapping classes based on hybrid neural networks], *Trudy 2-y Rossiyskoy konferentsii molodykh uchenykh po matematicheskoy modelirovaniyu* [Proceedings of 2-nd Russian conference of young scientists on mathematical modeling]. Moscow: MGTU, 2002, pp. 24-30.
12. *Krug P.G.* Neyronnye seti i neyrokomp'yutery: uchebnoe posobie po kursu «Mikroprotsessory» [Neural networks and Neurocomputers: textbook for the course "Microprocessors"]. Moscow: Izd-vo MEI, 2002, 176 p.

13. *Bregman L.M.* Dokazatel'stvo skhodimosti metoda Sheleykhovskogo dlya zadachi s transportnymi ogranicheniyami [The proof of convergence of the method Celakovskoho for the problem with transport constraints], *Vychislitel'naya matematika i matematicheskaya fizika* [Computational mathematics and mathematical physics]. Moscow, 1967, pp. 147-156.
14. *Grishkov A.F., Kulakov A.A., Shpak V.F.* Klassifikatsiya obrazov radiotekhnicheskikh signalov [Classification of images of radio signals], *Voprosy spetsial'noy radioelektroniki* [Questions of Special Radio Electronics], 2009, Issue 2, pp. 30-37.
15. *Trukhaev R.I.* Modeli prinyatiya resheniy v usloviyakh neopredelennostey [Models of decision making under conditions of uncertainty]. Moscow: Nauka, 1981, 258 p.
16. *Jaynes, E.T.* Information Theory and Statistical Mechanics, *Physical Review*, 1957, No. 106 (4), pp. 620-630. DOI:10.1103/PhysRev.106.620. Bibcode: 1957PhRv..106..620J.
17. *Jaynes E.T.* Information Theory and Statistical Mechanics II, *Physical Review*, 1957, No. 108 (2), pp. 171-190. DOI: 10.1103/PhysRev.108.171. Bibcode: 1957PhRv..108..171J.
18. *Shpak B.F., Getmanchuk A.V., Radchenko S.A., Kulakov A.A.* K voprosu o skhodimosti algoritma klassifikatsii radiotekhnicheskikh obrazov po metodu G.V. Sheleykhovskogo [To the question of convergence of the algorithm for the classification of electronic images by the method of G. V. Celakovskoho], *Voprosy spetsial'noy radioelektroniki* [Questions of Special Radio Electronics], 2011, Issue 2, pp. 41-46.
19. *Sinkhorn, Richard.* A relationship between arbitrary positive matrices and doubly stochastic matrices. *Ann. Math. Statist.*, 1964.
20. *Sinkhorn, Richard, & Knopp, Paul.* Concerning nonnegative matrices and doubly stochastic matrices. *Pacific J. Math.*, 1967.

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

Гетманчук Алексей Владимирович – Южный федеральный университет; e-mail: Getmanchuk@mail.ru; 347928, г. Таганрог, пер. Некрасовский, 44; тел.: 88634371673; кафедра математического обеспечения и применения ЭВМ; аспирант.

Getmanchuk Aleksei Vladimirovich – Southern Federal University; e-mail: Getmanchuk@mail.ru; 44, Nekrasovskiy, Taganrog, 347928, Russia; phone: +78634371651; the department of software engineering; postgraduate student.