

УДК 519.2:303.732.4

UDC 519.2:303.732.4

МАТЕМАТИЧЕСКИЕ МЕТОДЫ ТЕОРИИ КЛАССИФИКАЦИИ

MATHEMATICAL METHODS OF CLASSIFICATION THEORY

Орлов Александр Иванович
д.э.н., д.т.н., к.ф.-м.н., профессор

Orlov Alexander Ivanovich
Dr.Sci.Econ., Dr.Sci.Tech., Cand.Phys-Math.Sci.,
professor
Bauman Moscow State Technical University,
Moscow, Russia

*Московский государственный технический
университет им. Н.Э. Баумана, Россия, 105005,
Москва, 2-я Бауманская ул., 5, prof-orlov@mail.ru*

Дан обзор математических методов построения и использования классификаций. Рассмотрены основные подходы к решению задач кластер-анализа и группировки. Предложены глобальные и локальные критерии естественности классификации. Методы дискриминантного анализа (диагностики, распознавания образов с учителем) обсуждаются в связи с проблемами построения обобщенных показателей (рейтингов).

This article gives a review of mathematical methods of construction and using of classifications. The main approaches to solving the problems of cluster analysis and grouping are discussed. We have also proposed global and local natural classification criteria. The methods of discriminant analysis (diagnosis, pattern recognition with the teacher) are discussed in connection with the construction of generalized indicators (ratings)

Ключевые слова: МАТЕМАТИЧЕСКИЕ МЕТОДЫ ДИАГНОСТИКИ, ДИСКРИМИНАНТНЫЙ АНАЛИЗ, РАСПОЗНАВАНИЕ ОБРАЗОВ, КЛАСТЕР-АНАЛИЗ, ЕСТЕСТВЕННОСТЬ КЛАССИФИКАЦИИ, ПРОГНОСТИЧЕСКАЯ СИЛА, РЕЙТИНГИ, ТЕОРИЯ КЛАССИФИКАЦИИ

Keywords: MATHEMATICAL METHODS FOR DIAGNOSIS, DISCRIMINANT ANALYSIS, PATTERN RECOGNITION, CLUSTER ANALYSIS, NATURAL CLASSIFICATION, PREDICTIVE POWER, RATING, CLASSIFICATION THEORY

1. Введение. Основные понятия

Термин «классификация» имеет несколько основных смысла. Во-первых, это система классов. Во-вторых, это действие, связанное с системой классов. Согласно [1, с.6] «термином «классификация» обозначают, по крайней мере, три разные вещи: процедуру построения классификации (выделение классов – А.О.), построенную классификацию (систему выделенных классов – А.О.) и процедуру ее использования (например, правила отнесения вновь поступающего объекта к одному из ранее выделенных классов –А.О.)» Выделим естественную триаду: построение классификаций – их изучение – и применение, в соответствии с которой упорядочим анализ задач классификации.

Математическая теория классификации – обширная область прикладной статистики и эконометрики [2, 3]. Какие научные исследования относить к этой теории? Исходя из потребностей специалиста, применяющего математические методы классификации, целесообразно принять, что сюда входят исследования, во-первых, отнесенные самими авторами к этой теории; во-вторых, связанные с ней общностью тематики, хотя бы их авторы и не упоминали термин «классификация». Это предполагает сложную внутреннюю структуру рассматриваемой научной области.

Следует иметь в виду, что в литературе наряду с термином «классификация» в близких смыслах используются термины «группировка», «распознавание образов», «диагностика», «дискриминация», «сортировка», «типология», «таксономия» и др. Терминологический разнобой связан прежде всего с традициями научных кланов, к которым относятся авторы публикаций, а также с внутренним делением самой теории классификации.

В научных исследованиях по современной теории классификации можно выделить два относительно самостоятельных направления. Одно из них опирается на опыт таких наук, как биология, география, геология, и таких прикладных областей, как ведение классификаторов продукции и библиотечное дело. Типичные объекты рассмотрения - классификация химических элементов (таблица Д.И. Менделеева), биологическая систематика, универсальная десятичная классификация публикаций (УДК), классификатор товаров на основе штрих-кодов. Опыт этого направления с гносеологических позиций обобщен в [1], соответствующий математический аппарат приведен в [4, 5].

Другое направление опирается на опыт технических исследований, экономики, маркетинговых исследований, социологии, медицины. Типичные задачи - техническая и медицинская диагностика. А также, например, разбиение на группы отраслей промышленности, тесно связанных между собой, выделение групп однородной продукции. Обычно используются такие термины, как «кластер-анализ», «распознавание образов» или «дискриминантный анализ». [6]. Краткое осмысление опыта и современного состояния этого направления дано в [7].

Это направление обычно опирается на математические модели; для проведения расчетов интенсивно используется ЭВМ. Однако относить его к математике столь же нецелесообразно, как астрономию или квантовую механику. Рассматриваемые математические модели можно и нужно изучать на формальном уровне, и такие исследования проводятся. Но направление в целом сконцентрировано на решении конкретных задач прикладных областей и вносит вклад в технические или экономические науки, медицину, социологию, но, как правило, математики не рассматривают полученные научные результаты как вклад в математику.

В 60-х годах XX века внутри прикладной статистики (в понимании этой науки, раскрытом в [2, 8, 9]) достаточно четко оформилась область, посвященная методам классификации. Несколько модифицируя формулировки М. Дж. Кендалла и А. Стьюарта 1966 г. (см. русский перевод [10, с.437]), в теории классификации выделим три подобласти: кластеризация (кластер-анализ) и группировка, статистический анализ классификаций, дискриминация (дискриминантный анализ). Опишем эти подобласти.

При кластеризации и группировке целью является выявление и выделение классов. Синонимы: построение классификации, распознавание образов без учителя, автоматическая классификация без учителя, типология, таксономия и др. Задача *кластер-анализа* состоит в выяснении по эмпирическим данным, насколько элементы «группируются» или распадаются на изолированные «скопления», «кластеры» (от *cluster* (англ.) - гроздь, скопление). Иными словами, задача - выявление естественного разбиения на классы, свободного от субъективизма исследователя, а цель - выделение групп однородных объектов, сходных между собой, при резком отличии этих групп друг от друга.

При *группировке*, наоборот, «мы хотим разбить элементы на группы независимо от того, естественны ли границы разбиения или нет» [10, с.437]. Цель по-прежнему состоит в выявлении групп однородных объектов, сходных между собой (как в кластер-анализе), однако «соседние» группы могут не иметь резких различий (в отличие от кластер-анализа). Границы между группами условны, не являются естественными, зависят от субъективизма исследователя. Аналогично при лесоустройстве проведение просек (границ участков) зависит от специалистов лесного ведомства, а не от свойств леса.

Задачи кластеризации и группировки принципиально различны, хотя для их решения могут применяться одни и те же алгоритмы. Важная для практической деятельности проблема состоит в том, чтобы понять, разрешима ли задача кластер-анализа для конкретных данных или возможна только их группировка, поскольку совокупность объектов достаточно однородна и не разбивается на резко разделяющиеся между собой кластеры.

Как правило, в математических задачах кластеризации и группировки основное - выбор метрики, расстояния между объектами, меры близости, сходства, различия. Хорошо известно, что для любого заданного разбиения объектов на группы и любого числа $\epsilon > 0$ можно указать метрику такую, что расстояния между объектами из одной группы будут меньше ϵ , а между объектами из разных групп - больше $1/\epsilon$. Тогда любой разумный алгоритм кластеризации даст именно заданное разбиение. Поэтому весьма важен выбор метрики, адекватной решаемой прикладной задаче. Некоторые подходы к выбору расстояния в задачах классификации рассмотрены в обзоре [11].

В *дискриминантном анализе* классы предполагаются заданными - плотностями вероятностей или обучающими выборками. Задача состоит в том, чтобы вновь поступающий объект отнести в один из этих классов. У понятия «дискриминация» имеется много синонимов: диагностика, распознавание образов с учителем, автоматическая классификация с учителем, статистическая классификация и т.д. Для построения системы

диагностических классов используют разнообразные методы кластерного анализа и группировки объектов.

Наименее известен второй член триады (отсутствующий у Кендалла и Стьюарта [10]) – изучение отношений эквивалентности, полученных в результате построения системы диагностических классов. Например, эксперты разбивают объекты экспертизы на группы схожих между собой. Ответ каждого из них – классификация (т.е. разбиение на классы исходного множества объектов экспертизы, в другой терминологии – отношение эквивалентности). Как построить итоговое мнение комиссии экспертов? Статистический анализ отношений эквивалентности – часть статистики бинарных отношений и тем самым – статистики объектов нечисловой природы [2, 12]. Помимо общих результатов этой области прикладной статистики, представляют интерес частные результаты, полученные специально для отношений эквивалентности [13].

Диагностика в узком смысле слова (процедура использования классификации, т.е. отнесения вновь поступающего объекта к одному из выделенных ранее классов) – предмет дискриминантного анализа. Отметим, что с точки зрения статистики объектов нечисловой природы дискриминантный анализ является частным случаем общей схемы регрессионного анализа, соответствующим ситуации, когда зависимая переменная принимает конечное число значений, а именно – номера классов, а вместо квадрата разности стоит функция потерь от неправильной классификации [2, 12]. Однако есть ряд специфических постановок, выделяющих задачи диагностики среди всех регрессионных задач.

2. Основные постановки задач построения классификаций

Процедуры построения диагностических правил делятся на вероятностные и детерминированные. К первым относятся задачи *расщепления смесей* [15-17]. В них предполагается, что распределение вновь поступающего случайного элемента является смесью вероятностных законов, соответствующих диагностическим классам. Как и при выборе степени полинома в регрессии [2, 18], при анализе данных о веществах и материалах встает вопрос оценки числа элементов смеси, т.е. числа диагностических классов. Нами изучены результаты применения обычно рекомендуемого критерия Уилкса для оценки числа элементов смеси. Оказалось [19], что оценка с помощью критерия Уилкса не является состоятельной, асимптотическое распределение этой оценки – геометрическое, как и в случае задач восстановления зависимости [2, 18]. Итак, продемонстрирована несостоятельность обычно используемых оценок. Для получения состоятельных оценок достаточно связать уровень значимости в критерии Уилкса с объемом выборки, как это предложено в работах [20, 21] для задач регрессии.

В рамках задач расщепления смесей важно обратить внимание на одно распространенное заблуждение. Иногда рекомендуют сначала построить систему диагностических классов, а потом в каждом диагностическом классе отдельно проводить регрессионный анализ (в классическом смысле) или применять иные методы многомерного статистического анализа. Однако необходимо подчеркнуть, что при этом нельзя опираться на вероятностную модель многомерного нормального распределения, так как распределения результатов наблюдений, попавших в определенный кластер, будут отнюдь не нормальным, а усеченным нормальным (усечение определяется границами кластера), как это подробнее разъяснено в статье [19].

Перейдем к детерминированному случаю. Как уже отмечалось, задачи построения системы диагностических классов целесообразно разбить на два типа: с четко разделенными кластерами (задачи кластер-анализа) и с условными границами, непрерывно переходящими друг в друга классами (задачи группировки). Такое деление полезно, хотя в обоих случаях могут применяться одинаковые алгоритмы [22 - 25].

Сколько существует алгоритмов построения системы диагностических правил? Иногда называют то или иное число. На самом же деле их бесконечно много.

Действительно, часто применяется т.н. агломеративный иерархический алгоритм «Дендрограмма», в котором вначале все элементы рассматриваются как отдельные кластеры, а затем на каждом шагу объединяются два наиболее близких кластера. Для работы «Дендрограммы» необходимо задать правило вычисления расстояния между кластерами. Оно вычисляется через расстояние $d(x,y)$ между элементами x и y . Поскольку $d^a(x,y)$ при $0 < a < 1$ также расстояние, то каждому значению степени a соответствует свой алгоритм.

Расстояние между кластерами рассчитывается как средняя связь (отсюда и еще название – алгоритм средней связи), т.е. как среднее арифметическое расстояний между парами объектов, один из которых входит в первый кластер, а другой – во второй. После ряда шагов все объекты объединяются вместе, и результат работы алгоритма представляет собой дерево последовательных объединений (в терминах теории графов), т.н. «Дендрограмму». Из нее можно выделить кластеры разными способами. Один подход – исходя из заданного числа кластеров. Другой – из соображений предметной области. Третий – исходя из устойчивости (если разбиение долго не менялось при возрастании порога объединения – значит, оно отражает реальность). И т.д.

Каким из бесконечного (континуального) семейства алгоритмов средней связи пользоваться при обработке данных? Дело осложняется тем, что практически в любом пространстве мер близости различных видов существует весьма много [11]. Именно в связи с обсуждаемой проблемой

следует указать [26] на принципиальное различие между кластер-анализом и задачами группировки.

Если классы реальны (в соответствии с определением, данным в [19]), естественны, существуют на самом деле, четко отделены друг от друга, то любой алгоритм кластер-анализа их выделит. Следовательно, *в качестве критерия естественности классификации следует рассматривать устойчивость относительно выбора алгоритма кластер-анализа.*

Проверить устойчивость можно, применив к данным несколько подходов, например, столь непохожие алгоритмы, как «ближнего соседа» и «дальнего соседа». Они отличаются от алгоритма средней связи способом введения расстояния между кластерами. В алгоритме «ближнего соседа» расстоянием между кластерами называется минимальное из расстояний между парами объектов, один из которых входит в первый кластер, а другой - во второй. А в алгоритме дальнего соседа расстоянием между кластерами называется максимальное из расстояний между парами объектов, один из которых входит в первый кластер, а другой - во второй.

Если полученные результаты содержательно близки, то они адекватны действительности. В противном случае следует предположить, что естественной классификации не существует, задача кластер-анализа не имеет решения, и можно проводить только группировку.

В более общем подходе агломеративные иерархические алгоритмы типа «Дендрограмма» при различных значениях параметра a , $0 < a < 1$, применяются для обработки одних и тех же реальных данных. Если при всех a получается практически одинаковое разбиение элементов на кластеры, т.е. результат работы алгоритма устойчив по отношению к изменению a (в смысле общей схемы устойчивости, предложенной в [27]), то имеем «естественную» классификацию. В противном случае результат зависит от субъективно выбранного исследователем параметра a , т.е. задача кластер-анализа неразрешима (предполагаем, что выбор a нельзя специально обосновать). Задача группировки в этой ситуации имеет много решений. Из них можно выбрать одно по дополнительным критериям.

Следовательно, получаем эвристический критерий: если решение задачи кластер-анализа существует, то оно находится с помощью любого алгоритма. Целесообразно использовать наиболее простой. Так, для классификации социально-психологических характеристик способных к математике школьников [28] мы использовали алгоритм [29]. На программирование и счет на ЭВМ ушло около полугода. Через несколько лет с помощью алгоритмов «ближайшего соседа» и «дальнего соседа» кластер-анализ был проведен вручную за 1,5 часа. С содержательной точки

зрения, полученные разбиения отличались мало¹. Поэтому есть основания считать, что с помощью этих алгоритмов действительно выявлена «реальная» структура данных.

3. Проблема поиска естественной классификации

Широко обсуждается проблема поиска естественной классификации (в отличие от искусственной). Приведенные в [1, 7, 30] высказывания дают представление о больших расхождениях в понимании «естественной классификации». Констатируем, что этот термин является нечетким (в смысле математической теории нечеткости [31, 32]), как, впрочем, и многие другие термины, как научно-технические, так и используемые в быденном языке. В книге [33] и статье [32] подробно обоснована нечеткость естественного языка и тот факт, что «мы мыслим нечетко». Однако это не слишком мешает решать производственные и жизненные проблемы. Кажущееся рациональным требование выработать сначала строгие определения, а потом развивать науку - невыполнимо. Следовать ему - значит отвлекать силы от реальных задач. При системном подходе (в интерпретации монографии [4]) к теории классификации становится ясно, что строгие определения можно надеяться получить на последних этапах построения теории. Мы же сейчас находимся в начале пути. Поэтому, не давая здесь определения понятию «естественная классификация», обсудим, как проверить на «естественность» классификацию (набор диагностических классов), полученную расчетным путем.

Можно выделить два критерия «естественности», по поводу которых имеется относительное согласие:

А. Естественная классификация должна быть реальной, соответствующей действительному миру, лишенной внесенного исследователем субъективизма;

Б. Естественная классификация должна быть важной или с научной точки зрения (давать возможность прогноза, предсказания новых свойств, сжатия информации и т.д.), или с практической.

Пусть классификация проводится на основе информации об объектах, представленной в виде матрицы «объект-признак» или матрицы попарных расстояний (мер близости или различия). Пусть алгоритм классификации дал разбиение на кластеры. Как можно получить доводы в пользу естественности этой классификации? Например, уверенность в том, что она - закон природы, может появиться только в результате длительного ее изучения и практического применения. Это соображение относится и к другим перечисленным в [1] критериям, а также к критерию Б (важности). Сосредоточимся на критерии А (реальности).

¹ Фактически анализировались иерархические деревья разбиений, поскольку все три алгоритма включали одномерные параметры, смысл которых — расстояние между объединяемыми на очередном шагу кластерами.

Понятие «реальности» кластера требует специального обсуждения (оно начато нами в работе [19]). Рассмотрим существо различий между понятиями «классификация» и «группировка». Пусть, к примеру, необходимо деревья, растущие в определенной местности, разбить на группы находящихся рядом друг с другом. Ясна интуитивная разница между несколькими отдельными рощами, далеко отстоящими друг от друга и разделенными полями, и сплошным лесом, разбитым просеками на квадраты с целью лесоустройства. Однако формально определить эту разницу столь же сложно, как определить понятие «куча зерен», чем занимались еще в Древней Греции [32].

Переформулируем сказанное в терминах «кластер-анализа» и «методов группировки». Выделенные с помощью первого подхода кластеры реальны, а потому могут рассматриваться как кандидаты в «естественные». Группировка дает «искусственные» классы, которые не могут быть «естественными».

Выборку из унимодального распределения можно, видимо, рассматривать как «естественный», «реальный» кластер. Применим к ней какой-либо алгоритм классификации («Форель», «ближнего соседа» и т.п. [34 – 37]). Он даст разбиение на классы, которые, разумеется, не являются «реальными», поскольку отражают прежде всего свойства алгоритма, а не исходных данных. Как отличить такую ситуацию от противоположной, когда имеются реальные кластеры и алгоритм классификации более или менее точно их выделяет? Как известно, «критерий истины – практика», но слишком много времени необходимо для применения подобного критерия. Поэтому представляет интерес критерий, оценивающий «реальность» выделяемых с помощью алгоритма классификации кластеров одновременно с его применением.

Такой показатель существует - это критерий устойчивости. Общая схема формулирования и изучения проблем устойчивости выводов в математических моделях реальных явлений и процессов предложена в [27]. В частности, поскольку значения признаков всегда измеряются с погрешностями, то «реальное» разбиение должно быть устойчиво (т.е. не меняться или меняться слабо) при малых отклонениях исходных данных. Алгоритмов классификации существует бесконечно много, и «реальное» разбиение должно быть устойчиво по отношению к переходу к другому алгоритму. Другими словами, если «реальное» разбиение на диагностические классы возможно, то оно находится с помощью *любого* алгоритма автоматической классификации. Следовательно, критерием естественности классификации может служить совпадение результатов работы двух достаточно различающихся алгоритмов, например «ближнего соседа» и «дальнего соседа».

Нами рассмотрены² два типа «глобальных» критериев «естественности классификации», касающихся разбиения в целом. «Локальные» критерии относятся к отдельным кластерам. Простейшая постановка такова: достаточно ли однородны два кластера (две совокупности) для их объединения? Если объединение возможно, то кластеры не являются «естественными». Преимущество этой постановки в том, что она допускает применение статистических критериев однородности двух выборок.

В одномерном случае (классификация по одному признаку) разработано большое число подобных критериев — Смирнова, омега-квадрат (Лемана-Розенблатта), Вилкоксона, Ван-дер-Вардена, Стьюдента и др. [39 - 42]. Имеются критерии и для многомерных данных [26, 43]. Для одного из видов объектов нечисловой природы [12] — люсианов — статистические методы выделения «реальных» кластеров развиты в [12, 44].

Как и для иных методов прикладной статистики, свойства алгоритмов кластер-анализа необходимо изучать на вероятностных моделях [17]. Это требование относится и к условиям естественного объединения двух кластеров. Для многомерных кластеров робастная процедура проверки допустимости объединения предложена в [45], непараметрическая — в [26].

Вероятностные постановки нужно применять, в частности, при перенесении результатов, полученных по выборке, на генеральную совокупность [46]. Вероятностная теория кластер-анализа и методов группировки различна для исходных данных в виде таблиц «объект × признак» и в виде матриц сходства. Для первых параметрическая теория — это «расщепление смесей» [15-17]. Непараметрическая теория основана на непараметрических оценках плотностей вероятностей и их мод [14, 15, 30, 48]. Основные результаты, связанные с непараметрическими оценками плотностей в произвольных пространствах, обсуждаются ниже.

Если исходные данные — матрица сходства $\|d(x,y)\|$, то необходимо признать, что законченной вероятностно-статистической теории пока нет. Подходы к ее построению обсуждались в [19]. Рассмотрим модель, позволяющую разработать расчетные методы. Предположим, что результаты наблюдений можно рассматривать как выборку из некоторого распределения с монотонно убывающей плотностью при увеличении расстояния от некоторого центра. Примененный к подобным данным какой-либо алгоритм кластер-анализа порождает некоторое разбиение. Ясно, что оно — чисто формальное, поскольку выделенным таксонам (кластерам) не соответствуют никакие «реальные» классы. Другими словами, задача кластер-анализа не имеет решения, а алгоритм дает лишь

² Программная реализация осуществлена в разработанных под научным руководством автора пакетах ППАНД и ДИСАН [38].

группировку. При обработке реальных данных вид плотности неизвестен. Проблема состоит в том, чтобы определить результат работы алгоритма (реальные кластеры или формальные группы)³.

Частный случай этой проблемы - проверка обоснованности объединения двух кластеров, которые мы рассматриваем как два множества объектов, а именно, $\{a_1, a_2, \dots, a_k\}$ и $\{b_1, b_2, \dots, b_m\}$. Пусть, например, используется алгоритм типа «Дендрограммы». Ряд авторов высказывали следующую идею. Пусть есть две совокупности мер близости: внутри кластеров $d(a_i, a_j)$, $1 \leq i < j \leq k$, $d(b_\alpha, b_\beta)$, $1 \leq \alpha < \beta \leq m$ и между кластерами $d(a_i, b_\alpha)$, $1 \leq i \leq k$, $1 \leq \alpha \leq m$. Эти совокупности предлагается рассматривать как независимые выборки и проверять гипотезу о совпадении их функций распределения. Если гипотеза не отвергается, объединение кластеров считается обоснованным; в противном случае - объединять нельзя, алгоритм прекращает работу. В [49] для проверки однородности использовался критерий Вилкоксона U , а в [50] – критерий типа ω^2 (Лемана-Розенблатта).

В рассматриваемом подходе есть две некорректности. Во-первых, меры близости не являются независимыми случайными величинами. Во-вторых, не учитывается, что объединяются не произвольные заранее фиксированные кластеры, а полученные в результате работы некоторого алгоритма, и их состав оказывается случайным [19, разд. 4]. От первой из этих некорректностей можно частично избавиться. А именно, в [51] в предположении независимости и одинаковой распределенности элементов произвольного пространства $a_1, a_2, \dots, a_k, b_1, b_2, \dots, b_m$ установлено, что при достаточно большом числе элементов кластеров зависимость мер близости не влияет на распределение статистики Вилкоксона U (сумма рангов элементов первой совокупности мер близости в объединении двух описанных выше совокупностей мер близости). На основе этой теоремы разработан алгоритм проверки статистической гипотезы, согласно которой объединение двух кластеров образует однородную совокупность. Если величина U слишком мала, статистическая гипотеза однородности (т.е. обоснованности объединения двух кластеров) отклоняется (на заданном уровне значимости), и возможность объединения отбрасывается (подробнее см. в [30]).

Что касается глобальных критериев, то для изучения устойчивости по отношению к малым отклонениям исходных данных естественно использовать метод статистических испытаний и проводить расчеты по «возмущенным» данным. Некоторые теоретические утверждения, касающиеся влияния «возмущений» на кластеры различных типов, получены в статье [19].

Итак, одна из основных проблем при построении классификации - проверка «реальности» кластера, его объективного существования

³ Подробнее см. [19].

независимо от расчетов исследователя. Эта проблема давно обсуждается специалистами различных областей (см., например, [52]). Отметим, что идея устойчивости как критерия «реальности» иногда реализуется неадекватно. Так, в статье [53] для алгоритмов типа «Дендрограмма» предлагается выделять разбиения, которым соответствуют наибольшие приращения расстояния между кластерами между очередными объединениями кластеров. Для данных [28] это предложение не дало полезных результатов - были получены различные разбиения: три алгоритма - три разбиения. И с теоретической точки зрения предложение [53] несостоятельно, что нетрудно показать.

Действительно, рассмотрим алгоритм «ближнего соседа», использующий меру близости $d(x,y)$, и однопараметрическое семейство алгоритмов с мерой близости $d^a(x,y)$, $a>0$, также являющихся алгоритмами «ближнего соседа». Тогда дендрограммы, полученные с помощью этих алгоритмов, совпадают при всех a , поскольку при их реализации происходит лишь сравнение мер близости между объектами. Другими словами, дендрограмма, полученная с помощью алгоритма «ближнего соседа», является адекватной в порядковой шкале (измерения меры близости $d(x,y)$), т.е. сохраняется при любом строго возрастающем преобразовании этой меры [27]. Однако выделенные по методу [53] «устойчивые разбиения» меняются. В частности, при достаточно большом a «наиболее объективным» по [53] будет, как нетрудно показать, разбиение на два кластера! Таким образом, разбиение, выдвинутое в [53] как «устойчивое», на самом деле оказывается весьма неустойчивым.

Несколько слов о вычислительной сходимости алгоритмов кластер-анализа, другими словами, об устойчивости кластеров в процессе вычислений. Алгоритмы кластер-анализа и группировки зачастую являются итерационными. Например, формулируется правило улучшения шаг за шагом решения задачи кластер-анализа, но момент остановки вычислений не обсуждается. Примером является известный алгоритм «Форель» [22], в котором улучшается положение центра кластера. В этом алгоритме на каждом шаге строится шар определенного заранее радиуса, выделяются элементы кластеризуемой совокупности, попадающие в этот шар, и новый центр кластера строится как центр тяжести выделенных элементов. При анализе алгоритма «Форель» возникает проблема: завершится ли процесс улучшения центра кластера через конечное число шагов или же он может быть бесконечным. Она получила название «проблема остановки». Для широкого класса так называемых «эталонных алгоритмов» проблема остановки решена: процесс улучшения остановится через конечное число шагов [19].

Отметим, что алгоритмы кластер-анализа могут быть модифицированы разнообразными способами. Например, описывая алгоритм «Форель» в стиле статистики объектов нечисловой природы

[12], заметим, что вычисление центра тяжести для совокупности многомерных точек – это нахождение эмпирического среднего для меры близости, равной квадрату евклидова расстояния. Если взять более естественную меру близости – само евклидово расстояние, то получим эталонный алгоритм кластер-анализа «Медиана», отличающийся от «Форели» тем, что новый центр строится не с помощью средних арифметических координат элементов, попавших в кластер, а с помощью медиан.

Проблема остановки возникает не только при построении системы классов. Она принципиально важна и при оценивании параметров вероятностных распределений методом максимального правдоподобия. Обычно не представляет большого труда выписать систему уравнений максимального правдоподобия и предложить решать ее каким-либо численным методом. Однако когда остановиться, сколько итераций сделать, какая точность оценивания будет при этом достигнута? Общий ответ, видимо, невозможно найти, но обычно нет ответа и для конкретных семейств. Именно поэтому не рекомендуем решать системы уравнений максимального правдоподобия, предлагая вместо них одношаговые оценки [54]. Эти оценки задаются конечными формулами, но асимптотически столь же хороши, как и оценки максимального правдоподобия.

Исходными данными для решения задач кластер-анализа могут быть не только матрицы попарных мер близости (мер различия, расстояний), но и матрицы типа «объект \times признак». Есть двойственность: классифицироваться могут как объекты, так и признаки. Меры близости между признаками могут быть рассчитаны на основе коэффициентов корреляции [55]. Один из первых методов кластер-анализа – метод корреляционных плеяд [56] – предназначен именно для классификации признаков.

Сводке, анализу и сопоставлению различных методов кластер-анализа посвящены монографии [57-60]. Выделим разрабатываемую в ИПУ РАН глубокую теорию «классификационного анализа данных» [61 - 64], имеющую разнообразные применения [65, 66]. Интересные подходы к типологии и классификации разработаны в социологических исследованиях [67 - 69]. Оригинальные методы классификации разработаны и с успехом применены А.Т. Фоменко для статистического анализа исторических текстов [70]. Микроагрегирование оказалось полезным для защиты конфиденциальных данных [71]. Отметим алгоритмы снижения размерности (многомерного шкалирования, визуализации), позволяющие придать наглядность процедурам выделения кластеров [72 - 76].

Ряд важных результатов в теории и практике классификации достигнут без интенсивного применения математических методов [77, 78].

Однако в настоящей статье ограничиваемся рассмотрением только математических методов классификации.

4. Основные постановки задач дискриминантного анализа

Перейдем к этапу применения диагностических правил, когда классы, к одному из которых нужно отнести вновь поступающий объект, уже выделены. Разработано большое количество методов принятия диагностических решений [79 - 83]. Одни из них излагаются согласно традициям математической статистики [43, 84 - 87]. Среди них выделим работы [88 - 92], развивающие дискриминантный анализ в ситуации, когда число неизвестных параметров растет пропорционально объему выборки (асимптотика растущей размерности А.Н. Колмогорова).

Другие авторы предпочитают использовать термин «распознавание образов» [93 - 98]. С точки зрения статистики объектов нечисловой природы речь идет о восстановлении зависимости, когда функция принимает номинальные значения – номера классов [12, 99]. Специфика прикладных областей отражается на процедурах диагностики [100 - 102].

Для решения задач диагностики в вероятностно-статистической постановке используют два подхода – параметрический и непараметрический. Первый из них обычно основан на использовании того или иного индекса (рейтинга) и сравнения его с порогом. Индекс может быть построен по статистическим данным, например, как в классическом линейном дискриминантном анализе Фишера [103]. Часто индекс представляет собой линейную функцию от характеристик, выбранных специалистами предметной области, коэффициенты которой подбирают эмпирически. Непараметрический подход связан с леммой Неймана-Пирсона в математической статистике и с теорией статистических решений. Он опирается на использование непараметрических оценок плотностей распределений вероятностей, описывающих диагностические классы.

Обсудим ситуацию подробнее. Методы диагностики, как и статистические методы в целом, делятся на параметрические [104, 105] и непараметрические [106, 107]. Первые основаны на предположении, что классы описываются распределениями из некоторых параметрических семейств. Обычно рассматривают многомерные нормальные распределения, при этом зачастую без обоснования принимают гипотезу о том, что ковариационные матрицы для различных классов совпадают. Именно в таких предположениях сформулирован классический дискриминантный анализ Фишера. Как известно, обычно не только нет теоретических оснований считать, что наблюдения извлечены из нормального распределения, но и проверка статистических гипотез согласия с нормальным законом дает отрицательный результат [2, 3]. Известно также, что по выборкам, объем которых не превосходит 50,

нельзя сделать обоснованный вывод о принадлежности к нормальному закону [108].

Поэтому более корректными, чем параметрические, являются непараметрические методы диагностики. Исходная идея таких методов основана на лемме Неймана-Пирсона, входящей в стандартный курс математической статистики. Согласно этой лемме *оптимальное решение* об отнесении вновь поступающего объекта (сигнала, наблюдения и др.) к одному из двух классов принимается на основе отношения плотностей $f(x)/g(x)$, где $f(x)$ - плотность распределения, соответствующая первому классу, а $g(x)$ - плотность распределения, соответствующая второму классу.

Если плотности распределения неизвестны, то применяют их непараметрические оценки, построенные по обучающим выборкам. Пусть обучающая выборка объектов из первого класса состоит из n элементов, а обучающая выборка для второго класса - из m объектов. Тогда рассчитывают значения непараметрических оценок плотностей $f_n(x)$ и $g_m(x)$ для первого и второго классов соответственно, а диагностическое решение принимают по их отношению. Таким образом, для решения задачи диагностики достаточно научиться строить непараметрические оценки плотности для выборок объектов произвольной природы.

Ряд видов непараметрических оценок плотности распределения в пространствах произвольной природы предложен и изучен в [109, 48, 30, 110, 111]. Методы построения таких оценок плотности подробно рассмотрены в литературе по прикладной статистике и эконометрике [1, 2, 12]. На их основе этих оценок могут быть построены непараметрические правила диагностики. Их достоинством является универсальность, возможность применения без необходимости обоснования трудно проверяемых условий (например, нормальности распределения характеристик объектов оценки). Недостатком является отсутствие явных формул. Кроме того, для построения непараметрического решающего правила нужны обучающие выборки.

С теоретической точки зрения использование непараметрических оценок плотности дает возможность построить асимптотически оптимальное решающее правило. Как показано в работе [112], при выборе алгоритма для практического использования необходимо учитывать его быстродействие, допустимые объемы обучающих выборок и другие характеристики, в результате вместо алгоритма на основе отношения непараметрических оценок плотностей наиболее приемлемым может оказаться иной алгоритм.

Параметрическая теория обычно опирается на использование для описания классов многомерных нормальных распределений. В линейном дискриминантном анализе Р. Фишера [103] дополнительно предполагается, что ковариационные матрицы двух классов совпадают. Тогда оптимальное

правило диагностики использует некоторый порог K . Вновь поступающий объект относят к первому классу, если $f(x_1, x_2, \dots, x_m) < K$, и ко второму, если $f(x_1, x_2, \dots, x_m) \geq K$, где обобщающий показатель (рейтинг)

$$f(x_1, x_2, \dots, x_m) = a_1x_1 + a_2x_2 + \dots + a_mx_m \quad (1)$$

есть линейная функция от единичных показателей (факторов) x_1, x_2, \dots, x_m . Таким образом, теория классификации тесно связана с теорией рейтингов.

О математических теориях рейтингов. При разработке управленческих решений с целью совместного учета и соизмерении различных факторов, частичного снятия неопределенности широко используются рейтинги. В частности, для сведения к однокритериальной постановке могут быть применены методы построения единого (интегрального) критерия (рейтинга). Термин «*рейтинг*» происходит от английского «to rate» (оценивать) и «rating» (оценка, оценивание). Оценка – это число, градация качественного признака (удовл., хор., отл.), реже – упорядочение (ранжировка) или математический объект иной природы. Согласно [113, с.8]: «В современном понимании *рейтинг* – это комплексная оценка состояния субъекта, которая позволяет отнести его к некоторым классу или категории». Имеется достаточное согласие в среде специалистов для того, чтобы использовать термин «рейтинг» как синоним термина «интегральный (единый, обобщенный, системный) критерий (оценка, показатель), позволяющий сравнивать объекты (субъекты) с интересующей пользователя (этим термином) точки зрения. В частности, рейтинги целесообразно использовать при целеполагании (для соизмерения целей).

Для описания многообразия различных видов рейтингов выделим три варианта постановок задач:

1. Непосредственная оценка.
2. Оценка с использованием обучающих выборок.
3. Оценка на основе системы показателей с весовыми коэффициентами.

Под непосредственной оценкой понимаем постановку задачи, в которой итоговый результат получается при непосредственной обработке некоторого множества оценок, без привлечения дополнительной информации о дополнительных объектах или об оценках (весах) дополнительных факторов (признаков). Пример - усреднение чисел. В этом, казалось бы, простейшем примере возникает ряд вопросов. Каким средним пользоваться – средним арифметическим или медианой? Или иными – средним геометрическим и т.п. Некоторые ответы дает теория измерений [2, 3]. При усреднении других видов ответов экспертов теория усложняется. Например, усреднение бинарных отношений может проводиться путем расчета медианы Кемени [12, 114]. При этом может варьироваться как вид меры близости (расстояние Кемени, или его квадрат, или аналог на основе коэффициента ранговой корреляции

Спирмена, или D -метрика и т.п.), так и множество бинарных отношений, по которому проводится минимизация [2].

Для оценки с использованием обучающих выборок применяют линейный дискриминантный анализ Р. Фишера, непараметрический дискриминантный анализ на основе использования непараметрических оценок плотностей в пространствах произвольной природы, а также иные методы распознавания образов с учителем, в том числе нейросетевые [115]. Традиционная процедура такова. Первый этап - построение системы показателей. Сначала составляют возможно более полный исходный перечень. Затем список показателей сокращают. Например, проводят кластер-анализ показателей, оставляя из каждого кластера по одному представителю. Отбор информативного подмножества признаков в дискриминантном анализе – самостоятельный раздел прикладной статистики. Следующий этап – непосредственное построение линейного рейтинга на основе отобранных показателей с помощью алгоритмов дискриминантного анализа Фишера.

При оценке на основе системы показателей с весовыми коэффициентами основные составляющие процедур - показатели (факторы), индексы и границы. Для построения системы показателей, обычно иерархической (единичные показатели – групповые – обобщенный) применяют экспертные методы и статистические (формальные) методы выделения информативного подмножества признаков. Способы усреднения (расчета индексов) при переходе от единичных показателей к групповым и от групповых к обобщенному выбирают на основе тех же принципов, что и при непосредственной оценке, но с использованием взвешенных средних. Веса задают либо непосредственно, либо косвенно – с помощью экспертных упорядочений, парных сравнений или обучающих выборок (экспертно-статистический метод).

Важный частный случай - *бинарные рейтинги* [116], когда рейтинговая оценка принимает два значения. Объект оценки с помощью бинарного рейтинга относится к одному из двух классов. Следовательно, теория бинарных рейтингов – часть дискриминантного анализа, имеющего целью отнесение объекта к одному из двух классов, заданными плотностями вероятностей или обучающими выборками.

Часто строят рейтинг в виде функции $f(x_1, x_2, \dots, x_m)$ от единичных показателей (факторов) x_1, x_2, \dots, x_m , а для принятия решения используют порог K . Принимают определенное решение, если $f(x_1, x_2, \dots, x_m) < K$, и альтернативное, если $f(x_1, x_2, \dots, x_m) \geq K$. В этом случае для принятия решения используется бинарный рейтинг вида $g(f(x_1, x_2, \dots, x_m))$, где функция g принимает два значения, а именно, $g(z) = 0$ при $z < K$ и $g(z) = 1$ при $z \geq K$.

На основе бинарных рейтингов можно сконструировать рейтинг с большим числом градаций. Пусть рейтинговая оценка h принимает одно из трех значений $A < B < C$. С ней можно связать два бинарных рейтинга p и q , таких, что для первого из них $p = 0$ при $h < C$ и $p = 1$ при $h = C$, для второго $q = 0$ при $h < B$ и $q = 1$ при $h \geq B$. Ясно, что $h = A$ тогда и только тогда, когда $p = q = 0$, и $h = C$ тогда и только тогда, когда $p = q = 1$, в то время как $h = B$ тогда и только тогда, когда $p = 0, q = 1$. Таким образом, использование рейтинга h с тремя возможными значениями эквивалентно использованию двух бинарных рейтингов p и q .

Популярны линейные рейтинги, заданные формулой (1). Коэффициенты a_1, a_2, \dots, a_m называют коэффициентами важности (весомости, значимости). Их определяют либо экспертным путем, либо по статистическим данным, используя обучающие выборки [117-119]. Глубокая теория качественной и количественной важности критериев развита В.В. Подиновским [120-122].

По одним и тем же данным могут быть построены различные рейтинги. Например, с помощью обучающих выборок можно построить непараметрический бинарный рейтинг (заданный алгоритмически) и линейный рейтинг (по Р.Фишеру). В той же прикладной задаче может оказаться полезным также и линейный рейтинг на основе экспертных оценок коэффициентов. Есть и иные методы, в частности, основанные на использовании нейросетей для диагностики и рейтингования [115].

О сравнении алгоритмов диагностики по результатам обработки реальных данных. В прикладных исследованиях применяют различные методы дискриминантного анализа, основанные на вероятностно-статистических моделях, а также с ними не связанные, т.е. эвристические, использующие детерминированные методы анализа данных. Независимо от «происхождения», каждый подобный алгоритм должен быть исследован как на параметрических и непараметрических вероятностно-статистических моделях порождения данных, так и на различных массивах реальных данных. Цель такого исследования - выбор наилучшего алгоритма в определенной области применения, включение его в стандартные программные продукты, методические материалы, учебные программы и пособия. Но для этого надо уметь сравнивать алгоритмы диагностики по качеству. Как это делать?

При построении информационно-исследовательской системы диагностики материалов (ИИСДМ) (прообраз - диалоговая система «АРМ материалововеда», разработанная нами для ВНИИ эластомерных материалов) возникает задача сравнения прогностических правил «по силе». Прогностическое правило - это алгоритм, позволяющий по характеристикам материала прогнозировать его свойства. Если прогноз дихотомичен («есть» или «нет»), то правило является алгоритмом диагностики, при котором материал относится к одному из двух классов.

Прогностические правила могут быть извлечены из научно-технической литературы и практики. обычно каждое из них формулируется в терминах небольшого числа признаков, но наборы признаков сильно меняются от правила к правилу. Поскольку в ИИСДМ должно фиксироваться лишь ограниченное число признаков, то возникает проблема их отбора. Естественно отбирать лишь те из них, которые входят в наборы, дающие наиболее «надежные» прогнозы. Для придания точного смысла термину «надежный» необходимо иметь способ сравнения алгоритмов диагностики по прогностической «силе».

Результаты обработки реальных данных с помощью некоторого алгоритма диагностики в рассматриваемом случае двух классов описываются долями правильной диагностики κ - в первом классе и λ - во втором, с учетом долей классов в объединенной совокупности $\pi_j, j = 1, 2, \pi_1 + \pi_2 = 1$.

Нередко [6] как показатель качества алгоритма диагностики (прогностической «силы») используют долю правильной диагностики $\mu = \pi_1 \kappa + \pi_2 \lambda$. Однако показатель μ определяется, в частности, через характеристики π_1, π_2 , частично заданные исследователем (например, на них влияет тактика отбора образцов для изучения). В аналогичной медицинской задаче величина μ оказалась больше для тривиального прогноза, согласно которому у всех больных течение заболевания будет благоприятно. Тривиальный прогноз сравнивался с алгоритмом выделения больных с прогнозируемым тяжелым течением заболевания. Он был разработан группы под руководством академика АН СССР И.М. Гельфанда. Применение этого алгоритма с медицинской точки зрения вполне оправдано [123]. Итак, по доле правильной классификации μ алгоритм группы И.М. Гельфанда оказался хуже тривиального - объявить всех больных легкими, не требующими специального наблюдения. Этот вывод очевидно нелеп. И причина появления нелепости вполне понятна. Хотя доля тяжелых больных невелика, но смертельные исходы сосредоточены именно в этой группе больных. Поэтому целесообразна гипердиагностика - рациональнее часть легких больных объявить тяжелыми, чем сделать ошибку в противоположную сторону.

Применение теории статистических решений [124, 125] требует знания потерь от ошибочной диагностики, а в большинстве научно-технических и экономических задач определить потери сложно. В частности, из-за необходимости оценивать человеческую жизнь в денежных единицах. По этическим соображениям это, на наш взгляд, недопустимо.

Для выявления информативного набора признаков целесообразно использовать *метод пересчета на модель линейного дискриминантного анализа* [30, 51], согласно которому статистической оценкой

«прогностической силы» δ является «эмпирическая прогностическая сила» $\delta^* = \Phi(d^*/2)$, $d^* = \Phi^{-1}(\kappa) + \Phi^{-1}(\lambda)$, где $\Phi(x)$ - функция стандартного нормального распределения с математическим ожиданием 0 и дисперсией 1, а $\Phi^{-1}(y)$ - обратная ей функция.

Если классы описываются выборками из многомерных нормальных совокупностей с одинаковыми матрицами ковариаций, а для классификации применяется классический линейный дискриминантный анализ Р.Фишера, то величина d^* представляет собой состоятельную статистическую оценку расстояния Махаланобиса между двумя рассматриваемыми совокупностями независимо от порогового значения, определяющего конкретное решающее правило. В общем случае показатель δ^* вводится как эвристический. Распределение δ^* является асимптотически нормальным, что позволяет строить доверительные интервалы для δ [2, 3].

Как проверить обоснованность пересчета на модель линейного дискриминантного анализа? Допустим, что классификация состоит в вычислении некоторого прогностического индекса y и сравнении его с заданным порогом c . Объект относят к первому классу, если $y \leq c$, ко второму, если $y > c$. Прогностический индекс – это обычно линейная функция (1) от характеристик рассматриваемых объектов. Возьмем два значения порога c_1 и c_2 . Если пересчет на модель линейного дискриминантного анализа обоснован, то, как можно показать, «прогностические силы» для обоих правил совпадают: $\delta(c_1) = \delta(c_2)$. Выполнение этого равенства можно проверить как статистическую гипотезу. Расчетные алгоритмы предложены в [30, 51] и включены в [2, 3].

Экспертно-статистический метод. Оценивание экспертами коэффициентов линейного рейтинга не всегда надежно. Особенно в ситуации, когда экспертов мало, а разброс мнений экспертов велик. Тогда представляется целесообразным не оценивать коэффициенты, а привлечь высококвалифицированных экспертов для глобальной оценки, т.е. оценки непосредственно обобщающего показателя (рейтинга) $Y = f(x_1, x_2, \dots, x_m) = a_1x_1 + a_2x_2 + \dots + a_mx_m$.

Предположим сначала, что рейтинговые оценки высококвалифицированных экспертов являются числовыми. Тогда в качестве данных, исходных для статистического анализа, имеем выборку $(Y_i; x_{1i}, x_{2i}, \dots, x_{mi})$, $i = 1, 2, \dots, n$, где n – число ответов высококвалифицированных экспертов, содержащих глобальные оценки рейтинга для n ситуаций. С точки зрения прикладной статистики имеем задачу линейного регрессионного анализа, которая решается стандартными методами (с помощью непараметрического метода наименьших квадратов [2, 3, 126]).

Нет необходимости обязательно требовать, чтобы оценки высококвалифицированных экспертов являлись числами. Можно ограничиться результатами парных сравнений или ранжировками. Ясно, что такого рода глобальные оценки гораздо легче получить, и они будут более надежными (исходя из ранее обоснованного общего утверждения, что нечисловые ответы более естественны для экспертов, чем числовые). Затем по глобальным экспертным оценкам для n ситуаций можно состоятельно оценить коэффициенты линейного рейтинга [118].

В настоящее время теория рейтингов продолжает бурно развиваться. Так, проблемам обоснованного выбора коэффициентов важности посвящены работы В.В. Подиновского [120 - 122]. Сравнительный анализ пяти традиционных и четырех относительно новых методов нахождения коэффициентов важности бинарных (т.е. принимающих два значения) факторов осуществлен И.Ф. Шахновым [119]. При этом исходной информацией служат экспертные оценки, имеющие качественный характер.

Очевидна связь теории рейтингов с современной весьма математизированной теорией полезности [117], поскольку рейтинговая оценка – частный случай функций полезности, используемой для упорядочения объектов экспертизы.

5. Заключительные замечания

Математические методы теории классификации используют различный математический аппарат и имеют разнообразные применения. Эта статья – обзор основных источников. Всего публикаций (монографий, статей, тезисов и трудов конференций и др.) – десятки тысяч. В 1980-х гг. вопросами теории классификации (классиологии) активно занималась Комиссия ВСНТО по классификации (председатель – член-корреспондент АН СССР Г.Б. Бокий), вовлекшая в свою деятельность около тысячи отечественных специалистов. Тем не менее теория классификации остается чуть ли не самой хаотичной областью математических методов исследования.

Описанные в настоящей статье подходы и результаты применялись нами при решении ряда прикладных задач, в частности, при сертификации продукции, в задачах химической безопасности биосферы и экологического страхования, в медицине труда. Соответствующие алгоритмы включены в состав автоматизированных рабочих мест «АРМ материаловед» и «АРМ математика» (разработаны для ВНИИ эластомерных материалов и изделий), АРМ «Математика в экспертизе» (АРМ «МАТЭК»), диалоговых систем анализа статистических данных ДИСАН и ППАНД. Рассмотренные в статье подходы и результаты в области математических методов классификации постоянно используются в учебном процессе, многие из них включены в учебники [2, 3, 12].

Отметим разработку теоретических основ и технологии адаптивного семантического анализа в поддержке принятия решений при управлении сложными системами на основе теории распознавания образов [127, 128].

В настоящее время теория классификации быстро развивается. Предлагаются новые непараметрические статистические оценки и критерии [129], в частности, оценки плотности в пространствах произвольной природы [130] и прогностическая сила как показатель качества алгоритма диагностики [131]. Разрабатываются диагностические правила на основе информационного рассогласования Кульбака-Лейблера [132, 133], позволяющие повысить надежность решения задач распознавания образов [134]. Предложены методы кластеризация спонтанных последовательностей [135, 136], применяемых для описания поведения потребителей в маркетинговых исследованиях. Разрабатываются математические методы классификации текстовых документов [137, 138]. Дальнейшее развитие получила концепция устойчивости кластера как критерия его естественности [139, 140]. Получены новые результаты в области системно-когнитивного анализа изображений [141]. Существенно продвинуто [142] распознавание объектов по тупиковым тесторам (см. [97]). Из работ прикладной направленности обратим внимание на программное обеспечение интеллектуального анализа спектральных данных [143], применении методов классификации при построении хронологии [144] и при выявлении закономерностей распознавания социальных категорий [145].

Состояние и перспективы математических методов теории классификации обсуждаются в [146, 147]. Большинство этих методов основано на использовании расстояний (показателей различия) между объектами, а потому входят в статистику объектов нечисловой природы [148]. Путь к построению обширного класса новых математических методов теории классификации открывает системная нечеткая интервальная математика [149]. Заслуживает дальнейшего развития метод когнитивной кластеризации, обоснованный в системно-когнитивном анализе и реализованный в его программном инструментарии – интеллектуальной системе «Эйдос» [150]. В этом методе, в частности, критерий сходства объектов кластеризации – не евклидово расстояние или его варианты, а интегральный критерий неметрической природы – «суммарное количество информации».

Литература

1. Розова С.С. Классификационная проблема в современной науке. - Новосибирск: Наука, 1986. – 224 с.
2. Орлов А.И. Прикладная статистика. - М.: Экзамен, 2006. - 672 с.
3. Орлов А.И. Эконометрика. – М.: Экзамен, 2004. - 576 с.
4. Шрейдер Ю.А., Шаров А.А. Системы и модели. - М.: Радио и связь, 1982. – 152 с.

5. Воронин Ю.А. Теория классифицирования и ее приложения. – Новосибирск: Наука, 1985. – 232 с.
6. Горелик А.Л., Скрипкин В.А. Методы распознавания. - М.: Высшая школа, 1984. – 208 с.
7. Орлов А.И. Заметки по теории классификации // Социология: методология, методы, математические модели. 1991. №2. С.28-50.
8. Орлов А.И. Современная прикладная статистика // Заводская лаборатория. Диагностика материалов. 1998. Т.64. №3. С. 52 - 60.
9. Орлов А.И. Высокие статистические технологии // Заводская лаборатория. Диагностика материалов. 2003. Т.69. №11. С.55-60.
10. Кендалл М.Дж., Стьюарт А. Многомерный статистический анализ и временные ряды. - М.: Наука, 1976. – 736 с.
11. Раушенбах Г.В. Меры близости и сходства в социологии // Анализ нечисловой информации в социологических исследованиях. - М.: Наука, 1985. С.169-203.
12. Орлов А.И. Организационно-экономическое моделирование. Часть 1. Нечисловая статистика. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2009. – 544 с.
13. Маамяги А.В. Некоторые задачи статистического анализа классификаций. - Таллинн: Изд-во АН ЭССР, 1982. – 24 с.
14. Орлов А.И. Некоторые неклассические постановки в регрессионном анализе и теории классификации // Программно-алгоритмическое обеспечение анализа данных в медико-биологических исследованиях. - М.: Наука, 1987. С.27-40.
15. Айвазян С.А., Бежаева З.И., Староверов О.В. Классификация многомерных наблюдений. - М.: Статистика, 1974. – 240 с.
16. Апраушева Н.Н., Раджабова М.Б. Классификация хлопка-сырца по статистическому алгоритму. - М.: Вычислительный центр РАН, 1990. – 17 с.
17. Апраушева Н.Н. Новый подход к обнаружению кластеров. - М.: Вычислительный центр РАН, 1993. – 48 с.
18. Орлов А.И. Об оценивании регрессионного полинома // Заводская лаборатория. Диагностика материалов. 1994. Т.60, №5. С.43-47.
19. Орлов А.И. Некоторые вероятностные вопросы теории классификации // Прикладная статистика. - М.: Наука, 1983. – С.166-179.
20. Орлов А.И. Оценка размерности модели в регрессии // Алгоритмическое и программное обеспечение прикладного статистического анализа. -М.: Наука, 1980. С.92-99.
21. Орлов А.И. Асимптотика некоторых оценок размерности модели в регрессии // Прикладная статистика. - М.: Наука, 1983. – С.260-265.
22. Загоруйко И.Г. Методы распознавания и их применение. - М.: Советское радио, 1972. - 208 с.
23. Дюрэн Б., Оделл П. Кластерный анализ: Пер. с англ. -М.: Статистика, 1977. -125 с.
24. Классификация и кластер / Под ред. Дж. Вэн Райзина.- М: Мир, 1980.- 389 с.
25. Жамбю М. Иерархический кластер-анализ и соответствия. - М: Финансы и статистика, 1988. - 342 с.
26. Орлов А.И. Некоторые вероятностные вопросы кластер-анализа // Общая биология. Новые данные исследований структуры и функций биологических систем. Доклады МОИП, 1985. - М.: Наука, 1987. - С.53-56.
27. Орлов А.И. Устойчивость в социально-экономических моделях. - М.: Наука, 1979. – 296 с.

28. Орлов А.И.; Гусейнов Г.А. Математические методы в изучении способных к математике школьников // Исследования по вероятностно-статистическому моделированию реальных систем. - М.: ЦЭМИ АН СССР, 1977. С.80-93.
29. Куперштох В.Л., Миркин Б.Г., Трофимов В.А. Сумма внутренних связей как показатель качества классификации // Автоматика и телемеханика. 1976. № 3. С.91-98.
30. Орлов А.И. Математические методы исследования и диагностика материалов (Обобщающая статья) // Заводская лаборатория. Диагностика материалов. 2003. Т.69. №3. С.53-64.
31. Заде Л.А. Размытые множества и их применение в распознавании образов и кластер-анализе // Классификация и кластер / Под ред. Дж. Вэн Райзина.- М: Мир, 1980.- С. 208-247.
32. Орлов А.И. Математика нечеткости // Наука и жизнь. 1982. № 7. С.60-67.
33. Налимов В.В. Спонтанность сознания. Вероятностная теория смыслов и смысловая архитектура личности. – М.: Прометей, 1989. – 287 с.
34. Дорофеюк А.А. Алгоритмы автоматической классификации: Обзор // Автоматика и телемеханика. 1971. № 12. - С. 78 – 113.
35. Группировки и корреляции в экономико-статистических исследованиях. (Серия «Ученые записки по статистике», т.43.) - М.: Наука, 1982. – 373 с.
36. Загоруйко Н.Г. Прикладные методы анализа данных и знаний. — Новосибирск: ИМ СО РАН, 1999. — 270 с.
37. Борисова И.А., Загоруйко Н.Г., Кутненко О.А. Критерии информативности и пригодности подмножества признаков, основанные на функции сходства // Заводская лаборатория. Диагностика материалов. 2008. Т.74. №1. С. 68 – 71.
38. Легостаева И.Л., Орлов А.И. и др. Пакет программ анализа данных «ППАНД». - М.: Сотрудничающий центр ВОЗ по профессиональной гигиене, 1990. - 93 с.
39. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. - М.: Наука, 1983. – 416 с.
40. Орлов А.И. Какие гипотезы можно проверять с помощью двухвыборочного критерия Вилкоксона? // Заводская лаборатория. Диагностика материалов. 1999. Т.65. №1. С.51-55.
41. Кудлаев Э.М., Лагутин М.Б. Селекция параметрического семейства распределений // Заводская лаборатория. Диагностика материалов. 1999. Т.65. №5. С.54 – 59.
42. Орлов А.И. О проверке однородности двух независимых выборок // Заводская лаборатория. Диагностика материалов. 2003. Т.69. №1. С.55-60.
43. Андерсон Т. Введение в многомерный статистический анализ. - М.: ГИФМЛ, 1963. - 500 с.
44. Орлов А.И. Парные сравнения в асимптотике Колмогорова // Экспертные оценки в задачах управления. - М.: ИПУ, 1982. С.58 - 66.
45. Шурыгин А.М. Статистический кластер-критерий // Алгоритмическое и программное обеспечение прикладного статистического анализа. - М.: Наука, 1980. С.360-366.
46. Рекомендации. Прикладная статистика. Методы обработки данных. Основные требования и характеристики. - М.: ВНИИСИ, 1987. – 62 с.
47. Пярна К.А. Оптимальное разбиение метрического вероятностного пространства: Автореф. дисс. канд. физ.-мат. наук. Вильнюс, 1987. – 20 с.

48. Орлов А.И. Классификация объектов нечисловой природы на основе непараметрических оценок плотности // Проблемы компьютерного анализа данных и моделирования. - Минск: Белорусский государственный университет, 1991. С.141-148.
49. Райская Н.Н., Гостилин Н.Л., Френкель А.А. Об одном способе проверки обоснованности разбиения в кластерном анализе // Тезисы докладов Всесоюзной конференции «Применение многомерного статистического анализа в экономике и оценке качества продукции». – Тарту: Тартуский государственный университет, 1977. С.177-179.
50. Бала Ю.М., Фуки В.Б., Рог А.И. и др. О возможности автоматизации процесса дифференциальной диагностики атеросклеротического кардиосклероза и ревматических пороков сердца, осложненных мерцательной аритмией // Кардиология, 1977. Т.17. №7. С.55-61.
51. Орлов А.И. О сравнении алгоритмов классификации по результатам обработки реальных данных // Общая биология. Новые данные исследований структуры и функций биологических систем. Доклады МОИП, 1985. - М.: Наука, 1987. С.79-82.
52. Любичев А.А. Проблемы формы, систематики и эволюции организмов. - М.: Наука, 1982. – 278 с.
53. Плоткин А.А. Устойчивость разбиения как критерий оптимальности построенной классификации // Статистические методы анализа экспертных оценок. - М.: Наука, 1977. С.111-123.
54. Орлов А.И. О нецелесообразности использования итеративных процедур нахождения оценок максимального правдоподобия // Заводская лаборатория. Диагностика материалов. 1986. Т.52. № 5. С.67-69.
55. Кендэл М. Ранговые корреляции. - М.: Статистика, 1975. – 216 с.
56. Терентьев П.В. Метод корреляционных плеяд // Вестник ЛГУ. 1959. Т.9. Вып.2. С.137 – 144.
57. Бонгард М.М. Проблема узнавания. – М.: Наука, 1967. – 320 с.
58. Елисеева И.И., Рукавишников В.О. Группировка, корреляция, распознавание образов. – М.: Статистика, 1977. – 144 с.
59. Миркин Б.Г. Группировки в социально-экономических исследованиях: Методы построения и анализа. - М.: Финансы и статистика, 1985. – 223 с.
60. Мандель И.Д. Кластерный анализ. – М.: Финансы и статистика, 1988. – 176 с.
61. Бауман Е.В., Дорофеюк А.А. Классификационный анализ данных // Избранные труды Международной конференции по проблемам управления (МКПУ-99). Том 1. – М.: СИНТЕГ. 1999. - С. 62-67.
62. Бауман Е.В., Блудян Н.О. Методы нахождения глобальных экстремумов функционалов в задаче классификационного анализа данных // Труды Института проблем управления РАН, том XIII, М., ИПУ РАН, 2001, с. 129-136.
63. Чернявский А.Л., Бауман Е.В., Дорофеюк А.А. Методы динамического классификационного анализа данных // Искусственный интеллект. 2002. № 2. С.290-298.
64. Бауман Е.В., Москаленко Н.Е. Методы экстремальной группировки параметров долевого типа // Автоматика и телемеханика. 2008. №11. С.162 - 170.
65. Дорофеюк А.А. Методология экспертно-классификационного анализа в задачах управления и обработки сложноорганизованных данных (история и перспективы развития) // Проблемы управления. 2009. № 3.1. С. 19-28.
66. Бауман Е.В., Дорофеюк А.А., Никитина Т.А. Алгоритмы размытой упорядоченной классификации в задачах исследования социально-экономических

систем управления // Материалы Четвертой Международной конференции «Управление развитием крупномасштабных систем» MLSD'2010. - М.: ИПУ РАН, 2010. Т.1. С. 211-213.

67. Типология и классификация в социологических исследованиях. – М.: Наука, 1982. – 296 с.

68. Татарова Г.Г. Методология анализа данных в социологии. – М.: NOTA BENE, 1999. – 224 с.

69. Толстова Ю.Н. Анализ социологических данных. – М.: Научный мир, 2000. – 352 с.

70. Фоменко А.Т. Методы статистического анализа исторических текстов. Приложения к хронологии. Тт. I, II. – М.: КРАФТ+ЛЕАН, 1999. – 832 с. + 832 с.

71. Андреев-Андриевский А.Е. Применение микроагрегирования для защиты конфиденциальных данных // Заводская лаборатория. Диагностика материалов. 1998. Т.64. №1. С.51 – 55.

72. Перекрест В.Т. Нелинейный типологический анализ социально-экономической информации: Математические и вычислительные методы. - Л.: Наука, 1983. – 176 с.

73. Терехина А.Ю. Анализ данных методами многомерного шкалирования. - М.- Наука, 1986. – 168 с.

74. Айвазян С. А., Бухштабер В. М., Енюков И. С., Мешалкин Л. Д. Прикладная статистика: классификация и снижение размерности. — М.: Финансы и статистика, 1989. – 607 с.

75. Лагутин М.Б. Визуальное представление тесноты связей // Заводская лаборатория. Диагностика материалов. 2005. Т.71. №7. С.53 – 57.

76. Толстова Ю.Н. Основы многомерного шкалирования. – М.: КДУ, 2006. - 160 с.

77. Классификация и кодирование технико-экономической информации. – М.: Изд-во стандартов, 1972. – 176 с.

78. Бреховских С.М. Основы функциональной системологии материальных объектов. – М.: Наука. 1986. – 192 с.

79. Айзерман М.А., Браверман Э.М., Розоноэр Л.И. Метод потенциальных функций в теории обучения машин. – М.: Наука, 1970. – 384 с.

80. Дуда Р., Харт П. Распознавание образов и анализ сцен. — М.: Мир, 1976. – 511 с.

81. Шорников Б.С. Классификация и диагностика в биологическом эксперименте. – М.: Наука, 1979. – 141 с.

82. Duda R.O., Hart P.E., Stork D.H. Pattern Classification (2nd ed.). - Wiley Interscience. — 2000. – 738 p.

83. Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning.— Springer, 2001. — 533 p.

84. Статистические методы классификации. / Под ред. Ю.Н. Благовещенского. – М.: МГУ им. М.В.Ломоносова, 1969. – 174 с.

85. Абусев Р.А., Лумельский Я.П. Статистическая групповая классификация. – Пермь: Пермский государственный университет, 1987. – 92 с.

86. Абусев Р.А. Групповая классификация: Решающие правила и их характеристики. – Пермь: Пермский государственный университет, 1992. – 219 с.

87. Малиновский Л.Г. Модельно-структурные методы дискриминантного анализа в автоматизации задач классификации и прогноза. Автореф. дисс. ... докт. техн. наук. – М.: 1998. – 38 с.

88. Деев А.Д. Асимптотические разложения распределений статистик дискриминантного анализа // Статистические методы классификации. Вып.2. – М.: МГУ им. М.В. Ломоносова, 1972. – С.6-51.
89. Мешалкин Л.Д., Сердобольский В.И. Ошибки при классификации многомерных наблюдений // Теория вероятностей и ее применения. 1978. Т.23. №4. С.722 – 781.
90. Заруцкий В.И. Классификация нормальных векторов простой структуры в пространстве большой размерности // Прикладной многомерный статистический анализ. – М.: Наука, 1978. – С.37 – 51.
91. Сердобольский В.И., Орлов А.И. Статистический анализ при большом числе параметров // Тезисы докладов III Всесоюзной школы-семинара «Программно-алгоритмическое обеспечение прикладного многомерного статистического анализа». – М.: ЦЭМИ АН СССР, 1987. С. 151-160.
92. Гирко В.Л. Асимптотика распределения спектра случайных матриц // Успехи математических наук. 1989. Т.44. Вып.4. С.7-34.
93. Вапник В. Н., Червоненкис А. Я. Теория распознавания образов. — М.: Наука, 1974. — 416 с.
94. Левит В.Е., Переверзев-Орлов В.С. Структура и поле данных при распознавании образов. – М.: Наука, 1984. – 120 с.
95. Горелик А.Л., Гуревич И.Б., Скрипкин В.А. Современное состояние проблемы распознавания. – М.: Радио и связь, 1985. – 160 с.
96. Распознавание образов: Состояние и перспективы / Верхаген К., Дейн Р., Грун Ф. и др. / Пер. с англ. под ред. И.Б.Гуревича. – М.: Радио и связь, 1985. – 104 с.
97. Журавлёв Ю.И., Рязанов В.В., Сенько О.В. «Распознавание». Математические методы. Программная система. Практические применения. — М.: ФАЗИС, 2006. — 176 с.
98. Городецкий В.И., Серебряков С.В. Методы и алгоритмы коллективного распознавания // Автоматика и телемеханика. - 2008. - №11. - С.3 - 40.
99. Вапник В.Н. Восстановление зависимостей по эмпирическим данным. — М.: Наука, 1979. — 448 с.
100. Гуда А.Н. Математическое моделирование сложных технологических процессов железнодорожного транспорта. - Ростов н/Д.: Изд-во Рост. Ун-та, 1995. – 155 с.
101. Гуда А.Н. Методы анализа данных и принятия решений в затрудненных условиях. – Ростов н/Д.: Изд-во Северо-Кавказского научного центра высшей школы. 1997. – 139 с.
102. Лябах Н.Н., Шабельников А.Н. Техническая кибернетика на железнодорожном транспорте. - Ростов н/Д.: Изд-во Северо-Кавказского научного центра высшей школы. 2002. – 283 с.
103. Fisher R.A. The Use of Multiple Measurements in Taxonomic Problems // Ann. Eugenics. 1936. September. Vol.7. Pp.179-188. (Перевод: Фишер Рональд Э. Использование множественных измерений в задачах таксономии. – В сб.: Современные проблемы кибернетики. – М.: Знание, 1979. С.6-20.)
104. Малиновский Л.Г. Модельно-структурные методы в многомерном статистическом анализе (обобщающая статья) // Заводская лаборатория. Диагностика материалов. 2001. Т.67. №11. С.52 – 60.
105. Абусев Р.А. О групповом подходе в статистической классификации и контроле качества // Заводская лаборатория. Диагностика материалов. 2003. Т.69. №3. С.65 – 70.

106. Апраушева Н.Н., Гридина Е.Д. Дополнительные исследования по обнаружению кластеров. – М.: ВЦ РАН, 1991. – 20 с.
107. Штремель М.А., Кудря А.В., Иващенко А.В. Непараметрический дискриминантный анализ в задачах управления качеством // Заводская лаборатория. Диагностика материалов. 2006. Т.72. №5. С.53-62.
108. Селезнев В.Д., Денисов К.С. Исследование свойств критериев согласия функции распределения данных с гауссовой методом Монте-Карло для малых выборок // Заводская лаборатория. Диагностика материалов. . Т.71. №1. С.68 - 73.
109. Орлов А.И. Непараметрические оценки плотности в топологических пространствах // Прикладная статистика. - М.: Наука, 1983. С.12-40.
110. Богданов Ю.И. Информация Фишера и непараметрическая аппроксимация плотности распределения // Заводская лаборатория. Диагностика материалов. 1998. Т.64. №7. С.56 – 61.
111. Богданов Ю.И. Метод максимального правдоподобия и корневая оценка плотности распределения // Заводская лаборатория. Диагностика материалов. 2004. Т.70. №3. С.51 – 59.
112. Толчеев В.О. Модифицированный и обобщенный метод ближайшего соседа для классификации библиографических текстовых документов // Заводская лаборатория. Диагностика материалов. 2009. Т.75. №7. С.63-70.
113. Карминский А.М., Пересецкий А.А., Петров А.Е. Рейтинги в экономике: методология и практика. – М.: Финансы и статистика, 2005. – 240 с.
114. Кемени Дж., Снелл Дж. Кибернетическое моделирование: Некоторые приложения. - М.: Советское радио, 1972. – 192 с.
115. Корнеев Д.С. Использование аппарата нейронных сетей для создания модели оценки и управления рисками предприятия // Управление большими системами. Вып.17. – М.: ИПУ РАН, 2007. - С.81-102.
116. Орлов А.И. Бинарные рейтинги и их сравнение // Теория активных систем / Труды международной научно-практической конференции (14-15 ноября 2007 г., Москва, Россия). Общая редакция – В.Н. Бурков, Д.А. Новиков. – М.: ИПУ РАН, 2007. – С.186-190.
117. Фишберн П. Теория полезности для принятия решений. – М.: Наука, 1978. – 352 с.
118. Киселев Н.И. Экспертно-статистический метод определения функции предпочтения по результатам парных сравнений объектов // Алгоритмическое и программное обеспечение прикладного статистического анализа. – М.: Наука, 1980. – С.111-123.
119. Шахнов И.Ф. Некоторые модели квалиметрического анализа многофакторных объектов с бинарными факторами // Заводская лаборатория. Диагностика материалов. 2005. Т.71. №5. С.59-65.
120. Подиновский В.В. Анализ решений при множественных оценках коэффициентов важности критериев и вероятностей значений неопределенных факторов в целевой функции // Автоматика и телемеханика. 2004. №11. С.141-159.
121. Подиновский В.В. Количественная важность критериев с непрерывной шкалой первой порядковой метрики // Автоматика и телемеханика. 2005. №9. С.129-137.
122. Подиновский В.В. Введение в теорию важности критериев в многокритериальных задачах принятия решений. – М.: Физматлит, 2007. – 64 с.
123. Гельфанд И.М., Алексеевская М.А., Губерман Ш.А. и др. Прогнозирование исхода инфаркта миокарда с помощью программы «Кора-3» // Кардиология. 1977. Т.17. №6. С.19-23.

124. Блекуэлл Д., Гиршик М. Теория игр и статистических решений. - М.: ИЛ, 1958. -374 с.
125. Кини Р.Л., Райфа Х. Принятие решений при многих критериях: предпочтения и замещения. – М.: Радио и связь, 1981. – 560 с.
126. Орлов А.И. Восстановление зависимости методом наименьших квадратов на основе непараметрической модели с периодической составляющей // Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №07(091). С.189–218. – IDA [article ID]: 0911307013. – Режим доступа: <http://ej.kubagro.ru/2013/07/pdf/13.pdf>.
127. Луценко Е.В. Теоретические основы и технология адаптивного семантического анализа в поддержке принятия решений (На примере универсальной автоматизированной системы распознавания образов «ЭЙДОС-5.1») Монография (научное издание). – Краснодар: КЮИ МВД РФ, 1996. – 278 с.
128. Луценко Е.В., Симанков В.С. Адаптивное управление сложными системами на основе теории распознавания образов. Монография (научное издание). – Краснодар: ТУ КубГТУ, 1999. –318 с.
129. Орлов А.И., Толчеев В.О. Об использовании непараметрических статистических критериев для оценки точности методов классификации (обобщающая статья) // Заводская лаборатория. Диагностика материалов. 2011. Т.77. №3. С.58-66.
130. Орлов А.И. Оценки плотности в пространствах произвольной природы // Статистические методы оценивания и проверки гипотез: межвуз. сб. науч. тр. / Перм. гос. нац. иссл. ун-т. – Пермь, 2013. – Вып. 25. – С.21-33.
131. Орлов А.И. Прогностическая сила как показатель качества алгоритма диагностики. - В сб.: Статистические методы оценивания и проверки гипотез: межвуз. сб. науч. тр. Вып.23. – Пермь: Перм. гос. нац. иссл. ун-т, 2011. – С.104-116.
132. Савченко А.В. Метод направленного перебора альтернатив в задачах распознавания образов // Заводская лаборатория. Диагностика материалов. 2011. Т.77. №11. С.69-74.
133. Савченко А.В. Комбинирования диагностических правил на основе информационного рассогласования Кульбака-Лейблера // Заводская лаборатория. Диагностика материалов. 2012. Т.78. С.69-73.
134. Савченко А. В. Об одном способе повышения надежности решения задачи распознавания образов. // Заводская лаборатория. Диагностика материалов. 2013. Т.79. №1. С.72-77.
135. Горбач А.Н., Цейтлин Н.А. Кластеризация спонтанных последовательностей // Заводская лаборатория. Диагностика материалов. 2009. Т.75. №5. С.64-69.
136. Горбач А.Н., Цейтлин Н.А. Алгоритм кластеризации спонтанных последовательностей // Заводская лаборатория. Диагностика материалов. 2010. Т.76. №1. С.69-70.
137. Бородкин А.А., Толчеев В.О. Комплексная процедура редукции для увеличения быстродействия непараметрических методов классификации текстовых документов // Заводская лаборатория. Диагностика материалов. 2011. Т.77. №11. С.64-69.
138. Бородкин А.А., Толчеев В.О. Разработка и исследование методов взвешивания ближайших соседей (на примере классификации библиографических текстовых документов) // Заводская лаборатория. Диагностика материалов. 2013. Т.79. №7. С.70-74.
139. Орлов А.И. Устойчивость кластера – критерий его естественности // Математические методы изучения геологических явлений [Сб. ст.] Моск. об-во

испытателей природы, Межсекц. семинар по применению математики в геологии; [Гл. ред. А. Л. Яншин]. - М.: Наука/МОИП, 1990. – С. 54-60.

140. Орлов А.И. Устойчивость классификации относительно выбора метода кластер-анализа // Заводская лаборатория. Диагностика материалов. 2013. Т.79. №1. С.68-71.

141. Луценко Е.В. Системно-когнитивный анализ изображений (обобщение, абстрагирование, классификация и идентификация) // Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2009. – №02(46). – Шифр Информрегистра: 0420900012\0017. – Режим доступа: <http://ej.kubagro.ru/2009/02/pdf/10.pdf>

142. Макаркин А. А. Распознавание объектов по тупиковым тесторам // Заводская лаборатория. Диагностика материалов. 2013. Т.79. №3. С.63-71.

143. Баданов А.В., Борисова И.А., Дюбанов В.В., Загоруйко Н.Г., Кутненко О.А., Кучкин А.В., Мещеряков М.А., Милозворов Н.Г. Система СПЕКТРАН для интеллектуального анализа спектральных данных // Заводская лаборатория. Диагностика материалов. 2009. Т.75. №11. С.65-70.

144. Орлов А.И. Новая хронология как основа государственно-патриотического мировоззрения // Научный эксперт. – 2013. – №3. – С.76-87. – Режим доступа: http://problemanalysis.ru/text/Jornal3_2013.pdf.

145. Луценко Е.В., Трунев А.П. Фундаментальные закономерности распознавания социальных категорий по астрономическим данным на момент рождения // Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2008. – №10(44). – Шифр Информрегистра: 0420800012\0138. – Режим доступа: <http://ej.kubagro.ru/2008/10/pdf/01.pdf>

146. Орлов А.И. О развитии математических методов теории классификации // Заводская лаборатория. Диагностика материалов. 2009. Т.75. №7. С.51-63.

147. Новиков Д.А., Орлов А.И. Математические методы классификации // Заводская лаборатория. Диагностика материалов. 2012. Т.78. №4. С.3-5.

148. Орлов А.И. О развитии статистики объектов нечисловой природы / А.И. Орлов // Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №09(093). С. 273 – 309. – IDA [article ID]: 0931309019. – Режим доступа: <http://ej.kubagro.ru/2013/09/pdf/19.pdf>.

149. Орлов А.И., Луценко Е.В. Системная нечеткая интервальная математика (СНИМ) – перспективное направление теоретической и вычислительной математики // Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №07(091). С. 255 – 308. – IDA [article ID]: 0911307015. – Режим доступа: <http://ej.kubagro.ru/2013/07/pdf/15.pdf>.

150. Луценко Е.В., Коржаков В.Е. Метод когнитивной кластеризации или кластеризация на основе знаний (кластеризация в системно-когнитивном анализе и интеллектуальной системе «Эйдос») // Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2011. – №07(071). С. 528 – 576. – Шифр Информрегистра: 0421100012\0253, IDA [article ID]: 0711107040. – Режим доступа: <http://ej.kubagro.ru/2011/07/pdf/40.pdf>.

References

1. Rozova S.S. Klassifikacionnaja problema v sovremennoj nauke. - Novosibirsk: Nauka, 1986. – 224 s.
2. Orlov A.I. Prikladnaja statistika. - M.: Jekzamen, 2006. - 672 s.
3. Orlov A.I. Jekonometrika. – M.: Jekzamen, 2004. - 576 s.

4. Shrejder Ju.A., Sharov A.A. Sistemy i modeli. - M.: Radio i svjaz', 1982. – 152 s.
5. Voronin Ju.A. Teorija klassificirovanija i ee prilozhenija. – Novosibirsk: Nauka, 1985. – 232 s.
6. Gorelik A.L., Skripkin V.A. Metody raspoznavanija. - M.: Vysshaja shkola, 1984. – 208 s.
7. Orlov A.I. Zametki po teorii klassifikacii // Sociologija: metodologija, metody, matematicheskie modeli. 1991. №2. S.28-50.
8. Orlov A.I. Sovremennaja prikladnaja statistika // Zavodskaja laboratorija. Diagnostika materialov. 1998. T.64. №3. S. 52 - 60.
9. Orlov A.I. Vysokie statisticheskie tehnologii // Zavodskaja laboratorija. Diagnostika materialov. 2003. T.69. №11. S.55-60.
10. Kendall M.Dzh., St'juart A. Mnogomernyj statisticheskij analiz i vremennye rjady. - M.: Nauka, 1976. – 736 s.
11. Raushenbah G.V. Mery blizosti i shodstva v sociologii // Analiz nechislovoj informacii v sociologicheskikh issledovanijah. - M.: Nauka, 1985. S.169-203.
12. Orlov A.I. Organizacionno-jekonomicheskoe modelirovanie. Chast' 1. Nechislovaja statistika. – M.: Izd-vo MGTU im. N.Je. Bauman, 2009. – 544 s.
13. Maamjagi A.V. Nekotorye zadachi statisticheskogo analiza klassifikacij. - Tallinn: Izd-vo AN JeSSR, 1982. – 24 s.
14. Orlov A.I. Nekotorye neklassicheskie postanovki v regressionnom analize i teorii klassifikacii // Programmno-algoritmicheskoe obespechenie analiza dannyh v mediko-biologicheskikh issledovanijah. - M.: Nauka, 1987. S.27-40.
15. Ajvazjan S.A., Bezhaeva Z.I., Staroverov O.B. Klassifikacija mnogomernyh nabljudenij. - M.: Statistika, 1974. – 240 s.
16. Aprausheva N.N., Radzhabova M.B. Klassifikacija hlopka-syrca po statisticheskomu algoritmu. - M.: Vychislitel'nyj centr RAN, 1990. – 17 s.
17. Aprausheva N.N. Novyj podhod k obnaruzheniju klasterov. - M.: Vychislitel'nyj centr RAN, 1993. – 48 s.
18. Orlov A.I. Ob ocenivanii regressionnogo polinoma // Zavodskaja laboratorija. Diagnostika materialov. 1994. T.60, №5. S.43-47.
19. Orlov A.I. Nekotorye verojatnostnye voprosy teorii klassifikacii // Prikladnaja statistika. - M.: Nauka, 1983. – S.166-179.
20. Orlov A.I. Ocenka razmernosti modeli v regressii // Algoritmicheskoe i programmnoe obespechenie prikladnogo statisticheskogo analiza. -M.: Nauka, 1980. S.92-99.
21. Orlov A.I. Asimptotika nekotoryh ocenok razmernosti modeli v regressii // Prikladnaja statistika. - M.: Nauka, 1983. – S.260-265.
22. Zagorujko I.G. Metody raspoznavanija i ih primenenie. - M.: Sovetskoe radio, 1972. - 208 s.
23. Djuran B., Odell P. Klasternyj analiz: Per. s angl. -M.: Statistika, 1977. -125 s.
24. Klassifikacija i klaster / Pod red. Dzh. Vjen Rajzina.- M: Mir, 1980.- 389 s.
25. Zhambju M. Ierarhicheskij klaster-analiz i sootvetstvija. - M: Finansy i statistika, 1988. - 342 s.
26. Orlov A.I. Nekotorye verojatnostnye voprosy klaster-analiza // Obshhaja biologija. Novye dannye issledovanij struktury i funkcij biologicheskikh sistem. Doklady MOIP, 1985. - M.: Nauka, 1987. - S.53-56.
27. Orlov A.I. Ustojchivost' v social'no-jekonomicheskikh modeljah. - M.: Nauka, 1979. – 296 s.

28. Orlov A.I.; Gusejnov G.A. Matematicheskie metody v izuchenii sposobnyh k matematike shkol'nikov // Issledovaniya po veroyatnostno-statisticheskomu modelirovaniyu real'nyh sistem. - M.: CJeMI AN SSSR, 1977. S.80-93.
29. Kupershtoh B.JI., Mirkin B.G., Trofimov V.A. Summa vnutrennih svyazey kak pokazatel' kachestva klassifikacii // Avtomatika i telemekhanika. 1976. № 3. S.91-98.
30. Orlov A.I. Matematicheskie metody issledovaniya i diagnostika materialov (Obobshhajushhaja stat'ja) // Zavodskaja laboratorija. Diagnostika materialov. 2003. T.69. №3. S.53-64.
31. Zade L.A. Razmytye mnozhestva i ih primenenie v raspoznavanii obrazov i klaster-analize // Klassifikacija i klaster / Pod red. Dzh. Vjen Rajzina.- M: Mir, 1980.- S. 208-247.
32. Orlov A.I. Matematika nechetkosti // Nauka i zhizn'. 1982. № 7. S.60-67.
33. Nalimov V.V. Spontannost' soznaniya. Veroyatnostnaja teorija smyslov i smyslovaja arhitektonika lichnosti. - M.: Prometej, 1989. - 287 s.
34. Dorofejuk A.A. Algoritmy avtomaticheskoy klassifikacii: Obzor // Avtomatika i telemekhanika. 1971. № 12. - S. 78 – 113.
35. Gruppirovki i korrelyacii v jekonomiko-statisticheskikh issledovaniyah. (Serija «Uchenye zapiski po statistike», t.43.) - M.: Nauka, 1982. - 373 s.
36. Zagorujko N.G. Prikladnye metody analiza dannyh i znaniy. — Novosibirsk: IM SO RAN, 1999. — 270 s.
37. Borisova I.A., Zagorujko N.G., Kutnenko O.A. Kriterii informativnosti i prigodnosti podmnozhestva priznakov, osnovannye na funkcii shodstva // Zavodskaja laboratorija. Diagnostika materialov. 2008. T.74. №1. S. 68 – 71.
38. Legostaeva I.L., Orlov A.I. i dr. Paket programm analiza dannyh «PPAND». - M.: Sotrudnichajushhij centr VOZ po professional'noj gigijene, 1990. - 93 s.
39. Bol'shev L.N., Smirnov H.V. Tablicy matematicheskoy statistiki. - M.: Nauka, 1983. - 416 s.
40. Orlov A.I. Kakie gipotezy mozno proverjat' s pomoshh'ju dvuhvyborochnogo kriterija Vilkoksona? // Zavodskaja laboratorija. Diagnostika materialov. 1999. T.65. №1. S.51-55.
41. Kudlaev Je.M., Lagutin M.B. Selekcija parametricheskogo semejstva raspredelenij // Zavodskaja laboratorija. Diagnostika materialov. 1999. T.65. №5. S.54 – 59.
42. Orlov A.I. O proverke odnorodnosti dvuh nezavisimyh vyborok // Zavodskaja laboratorija. Diagnostika materialov. 2003. T.69. №1. S.55-60.
43. Anderson T. Vvedenie v mnogomernyj statisticheskij analiz. - M.: GIFML, 1963. - 500 s.
44. Orlov A.I. Parnye sravneniya v asimptotike Kolmogorova // Jekspertnye ocenki v zadachah upravleniya. - M.: IPU, 1982. S.58 - 66.
45. Shurygin A.M. Statisticheskij klaster-kriterij // Algoritmicheskoe i programmnoe obespechenie prikladnogo statisticheskogo analiza. - M.: Nauka, 1980. S.360-366.
46. Rekomendacii. Prikladnaja statistika. Metody obrabotki dannyh. Osnovnye trebovaniya i harakteristiki. - M.: VNIISI, 1987. - 62 s.
47. Pjarna K.A. Optimal'noe razbienie metricheskogo veroyatnostnogo prostranstva: Avtoref. diss. kand. fiz.-mat. nauk. Vil'njus, 1987. - 20 s.
48. Orlov A.I. Klassifikacija ob#ektov nechislovoj prirody na osnove neparametricheskikh ocenok plotnosti // Problemy komp'juternogo analiza dannyh i modelirovaniya. - Minsk: Belorusskij gosudarstvennyj universitet, 1991. S.141-148.
49. Rajskaia H.N., Gostilin N.L., Frenkel' A.A. Ob odnom sposobe proverki obosnovannosti razbieniya v klasterenom analize // Tezisy dokladov Vsesojuznoj konferencii

«Primenenie mnogomernogo statisticheskogo analiza v jekonomike i ocenke kachestva produkcii». – Tartu: Tartuskij gosudarstvennyj universitet, 1977. S.177-179.

50. Bala Ju.M., Fuki B.B., Rog A.I. i dr. O vozmozhnosti avtomatizacii processa differencial'noj diagnostiki ateroskleroticheskogo kardioskleroza i revmaticheskikh porokov serdca, oslozhnennykh mercatel'noj aritmiej // Kardiologija, 1977. T.17. №7. S.55-61.

51. Orlov A.I. O sravnenii algoritmov klassifikacii po rezul'tatam obrabotki real'nyh dannyh // Obshhaja biologija. Novye dannye issledovanij struktury i funkcij biologicheskikh sistem. Doklady MOIP, 1985. – M.: Nauka, 1987. S.79-82.

52. Ljubishhev A.A. Problemy formy, sistematiki i jevoljucii organizmov. – M.: Nauka, 1982. – 278 s.

53. Plotkin A.A. Ustojchivost' razbivenija kak kriterij optimal'nosti postroennoj klassifikacii // Statisticheskie metody analiza jekspertnyh ocenok. – M.: Nauka, 1977. S.111-123.

54. Orlov A.I. O necelesoobraznosti ispol'zovanija iterativnyh procedur nahozhdenija ocenok maksimal'nogo pravdopodobija // Zavodskaja laboratorija. Diagnostika materialov. 1986. T.52. № 5. S.67-69.

55. Kendjel M. Rangovyje korrelyacii. – M.: Statistika, 1975. – 216 s.

56. Terent'ev P.V. Metod korrelyacionnyh plejad // Vestnik LGU. 1959. T.9. Vyp.2. S.137 – 144.

57. Bongard M.M. Problema uznavanija. – M.: Nauka, 1967. – 320 s.

58. Eliseeva I.I., Rukavishnikov V.O. Gruppirovka, korrelyacija, raspoznavanie obrazov. – M.: Statistika, 1977. – 144 s.

59. Mirkin B.G. Gruppirovki v social'no-jekonomicheskikh issledovanijah: Metody postroenija i analiza. – M.: Finansy i statistika, 1985. – 223 s.

60. Mandel' I.D. Klasternyj analiz. – M.: Finansy i statistika, 1988. – 176 s.

61. Bauman E.V., Dorofejuk A.A. Klassifikacionnyj analiz dannyh // Izbrannye trudy Mezhdunarodnoj konferencii po problemam upravlenija (MKPU-99). Tom 1. – M.: SINTEG. 1999. – S. 62-67.

62. Bauman E.V., Bludjan N.O. Metody nahozhdenija global'nyh jekstremumov funkcionalov v zadache klassifikacionnogo analiza dannyh // Trudy Instituta problem upravlenija RAN, tom HIII, M., IPU RAN, 2001, s. 129-136.

63. Chernjavskij A.L., Bauman E.V., Dorofejuk A.A. Metody dinamicheskogo klassifikacionnogo analiza dannyh // Iskusstvennyj intellekt. 2002. № 2. C.290-298.

64. Bauman E.V., Moskalenko N.E. Metody jekstremal'noj gruppirovki parametrov dolevogo tipa // Avtomatika i telemekhanika. 2008. №11. S.162 - 170.

65. Dorofejuk A.A. Metodologija jekspertno-klassifikacionnogo analiza v zadachah upravlenija i obrabotki slozhnoorganizovannyh dannyh (istorija i perspektivy razvitija) // Problemy upravlenija. 2009. № 3.1. S. 19-28.

66. Bauman E.V., Dorofejuk A.A., Nikitina T.A. Algoritmy razmytoj uporyadochennoj klassifikacii v zadachah issledovanija social'no-jekonomicheskikh sistem upravlenija // Materialy Chetvertoj Mezhdunarodnoj konferencii «Upravlenie razvitiem krupnomasshtabnyh sistem» MLSD'2010. – M.: IPU RAN, 2010. T.1. S. 211-213.

67. Tipologija i klassifikacija v sociologicheskikh issledovanijah. – M.: Nauka, 1982. – 296 s.

68. Tatarova G.G. Metodologija analiza dannyh v sociologii. – M.: NOTA BENE, 1999. – 224 s.

69. Tolstova Ju.N. Analiz sociologicheskikh dannyh. – M.: Nauchnyj mir, 2000. – 352 s.

70. Fomenko A.T. Metody statisticheskogo analiza istoricheskikh tekstov. Prilozhenija k hronologii. Tt. I, II. – M.: KRAFT+LEAN, 1999. – 832 s. + 832 s.
71. Andreev-Andrievskij A.E. Primenenie mikroagregirovanija dlja zashhity konfidencial'nyh dannyh // Zavodskaja laboratorija. Diagnostika materialov. 1998. T.64. №1. S.51 – 55.
72. Perekrst V.T. Nelinejnyj tipologicheskij analiz social'no-jekonomicheskoy informacii: Matematicheskie i vychislitel'nye metody. - L.: Nauka, 1983. – 176 s.
73. Terehina A.Ju. Analiz dannyh metodami mnogomernogo shkalirovanija. - M.- Nauka, 1986. – 168 s.
74. Ajvazjan S. A., Buhshtaber V. M., Enjukov I. S., Meshalkin L. D. Prikladnaja statistika: klassifikacija i snizhenie razmernosti. — M.: Finansy i statistika, 1989. – 607 s.
75. Lagutin M.B. Vizual'noe predstavlenie tesnoty svyazej // Zavodskaja laboratorija. Diagnostika materialov. 2005. T.71. №7. S.53 – 57.
76. Tolstova Ju.N. Osnovy mnogomernogo shkalirovanija. – M.: KDU, 2006. - 160 s.
77. Klassifikacija i kodirovanie tehniko-jekonomicheskoy informacii. – M.: Izd-vo standartov, 1972. – 176 s.
78. Brehovskih S.M. Osnovy funkcional'noj sistemologii material'nyh ob#ektov. – M.: Nauka. 1986. – 192 s.
79. Ajzerman M.A., Braverman Je.M., Rozonojer L.I. Metod potencial'nyh funkcij v teorii obuchenija mashin. – M.: Nauka, 1970. – 384 s.
80. Duda R., Hart P. Raspoznavanie obrazov i analiz scen. — M.: Mir, 1976. – 511 s.
81. Shornikov B.S. Klassifikacija i diagnostika v biologicheskom jeksperimente. – M.: Nauka, 1979. – 141 s.
82. Duda R.O., Hart P.E., Stork D.H. Pattern Classification (2nd ed.). - Wiley Interscience. — 2000. – 738 p.
83. Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning.— Springer, 2001. — 533 p.
84. Statisticheskie metody klassifikacii. / Pod red. Ju.N. Blagoveshhenskogo. – M.: MGU im. M.V.Lomonosova, 1969. – 174 s.
85. Abusev R.A., Lumel'skij Ja.P. Statisticheskaja gruppovaja klassifikacija. – Perm': Permskij gosudarstvennyj universitet, 1987. – 92 s.
86. Abusev R.A. Gruppovaja klassifikacija: Reshajushhie pravila i ih harakteristiki. – Perm': Permskij gosudarstvennyj universitet, 1992. – 219 s.
87. Malinovskij L.G. Model'no-strukturnye metody diskriminantnogo analiza v avtomatizacii zadach klassifikacii i prognoza. Avtoref. diss. ... dokt. tehn. nauk. – M.: 1998. – 38 s.
88. Deev A.D. Asimptoticheskie razlozhenija raspredelenij statistik diskriminantnogo analiza // Statisticheskie metody klassifikacii. Vyp.2. – M.: MGU im. M.V. Lomonosova, 1972. - S.6-51.
89. Meshalkin L.D., Serdobol'skij V.I. Oshibki pri klassifikacii mnogomernyh nabljudenij // Teorija verojatnostej i ee primenenija. 1978. T.23. №4. S.722 - 781.
90. Zaruckij V.I. Klassifikacija normal'nyh vektorov prostoj struktury v prostranstve bol'shoj razmernosti // Prikladnoj mnogomernyj statisticheskij analiz. – M.: Nauka, 1978. – S.37 – 51.
91. Serdobol'skij V.I., Orlov A.I. Statisticheskij analiz pri bol'shom chisle parametrov // Tezisy dokladov III Vsesojuznoj shkoly-seminara «Programmno-algoritmicheskoe obespechenie prikladnogo mnogomernogo statisticheskogo analiza». - M.: CJeMI AN SSSR, 1987. S. 151-160.

92. Girko V.L. Asimptotika raspredelenija spektra sluchajnyh matric // Uspehi matematicheskikh nauk. 1989. T.44. Vyp.4. S.7-34.
93. Vapnik V. N., Chervonenkis A. Ja. Teorija raspoznavanija obrazov. — M.: Nauka, 1974. — 416 s.
94. Levit V.E., Pereverzev-Orlov V.S. Struktura i pole dannyh pri raspoznavanii obrazov. — M.: Nauka, 1984. — 120 s.
95. Gorelik A.L., Gurevich I.B., Skripkin V.A. Sovremennoe sostojanie problemy raspoznavanija. — M.: Radio i svjaz', 1985. — 160 s.
96. Raspoznavanie obrazov: Sostojanie i perspektivy / Verhagen K., Dejn R., Grun F. i dr. / Per. s angl. pod red. I.B.Gurevicha. — M.: Radio i svjaz', 1985. — 104 s.
97. Zhuravljov Ju.I., Rjazanov V.V., Sen'ko O.V. «Raspoznavanie». Matematicheskie metody. Programmaja sistema. Prakticheskie primenenija. — M.: FAZIS, 2006. — 176 s.
98. Gorodeckij V.I., Serebrjakov S.V. Metody i algoritmy kollektivnogo raspoznavanija // Avtomatika i telemekhanika. - 2008. - №11. - S.3 - 40.
99. Vapnik V.N. Vosstanovlenie zavisimostej po jempiricheskim dannym. — M.: Nauka, 1979. — 448 s.
100. Guda A.N. Matematicheskoe modelirovanie slozhnyh tehnologicheskikh processov zheleznodorozhnogo transporta. - Rostov n/D.: Izd-vo Rost. Un-ta, 1995. — 155 s.
101. Guda A.N. Metody analiza dannyh i prinjatija reshenij v zatrudnennyh uslovijah. — Rostov n/D.: Izd-vo Severo-Kavkazskogo nauchnogo centra vysshej shkoly. 1997. — 139 s.
102. Ljabah N.N., Shabel'nikov A.N. Tehnicheskaja kibernetika na zheleznodorozhnom transporte. - Rostov n/D.: Izd-vo Severo-Kavkazskogo nauchnogo centra vysshej shkoly. 2002. — 283 s.
103. Fisher R.A. The Use of Multiple Measurements in Taxonomic Problems // Ann. Eugenics. 1936. September. Vol.7. Pp.179-188. (Perevod: Fisher Ronal'd Je. Ispolzovanie mnozhestvennyh izmerenij v zadachah taksonomii. — V sb.: Sovremennye problemy kibernetiki. — M.: Znanie, 1979. S.6-20.)
104. Malinovskij L.G. Model'no-strukturnye metody v mnogomernom statisticheskom analize (obobshhajushhaja stat'ja) // Zavodskaja laboratorija. Diagnostika materialov. 2001. T.67. №11. S.52 — 60.
105. Abusev R.A. O gruppovom podhode v statisticheskoj klassifikacii i kontrole kachestva // Zavodskaja laboratorija. Diagnostika materialov. 2003. T.69. №3. S.65 — 70.
106. Aprausheva N.N., Gridina E.D. Dopolnitel'nye issledovanija po obnaruzheniju klasterov. — M.: VC RAN, 1991. — 20 s.
107. Shtremel' M.A., Kudrja A.V., Ivashhenko A.V. Neparametricheskij diskriminantnyj analiz v zadachah upravlenija kachestvom // Zavodskaja laboratorija. Diagnostika materialov. 2006. T.72. №5. S.53-62.
108. Seleznev V.D., Denisov K.S. Issledovanie svojstv kriteriev soglasija funkcii raspredelenija dannyh s gaussovoj metodom Monte-Karlo dlja malyh vyborok // Zavodskaja laboratorija. Diagnostika materialov. . T.71. №1. S.68 - 73.
109. Orlov A.I. Neparametricheskie ocenki plotnosti v topologicheskikh prostranstvah // Prikladnaja statistika. - M.: Nauka, 1983. S.12-40.
110. Bogdanov Ju.I. Informacija Fishera i neparametricheskaja approksimacija plotnosti raspredelenija // Zavodskaja laboratorija. Diagnostika materialov. 1998. T.64. №7. S.56 — 61.
111. Bogdanov Ju.I. Metod maksimal'nogo pravdopodobija i kornevaja ocenka plotnosti raspredelenija // Zavodskaja laboratorija. Diagnostika materialov. 2004. T.70. №3. S.51 — 59.

112. Tolcheev V.O. Modificirovannyj i obobshhennyj metod blizhajshego soseda dlja klassifikacii bibliograficheskikh tekstovyh dokumentov // Zavodskaja laboratorija. Diagnostika materialov. 2009. T.75. №7. S.63-70.
113. Karminskij A.M., Pereseckij A.A., Petrov A.E. Rejtingi v jekonomike: metodologija i praktika. – M.: Finansy i statistika, 2005. – 240 s.
114. Kemeni Dzh., Snell Dzh. Kiberneticheskoe modelirovanie: Nekotorye prilozhenija. - M.: Sovetskoe radio, 1972. – 192 s.
115. Korneev D.S. Ispol'zovanie apparata nejronnyh setej dlja sozdaniya modeli ocenki i upravlenija riskami predpriyatija // Upravlenie bol'shimi sistemami. Vyp.17. – M.: IPU RAN, 2007. - S.81-102.
116. Orlov A.I. Binarnye rejtingi i ih sravnenie // Teorija aktivnyh sistem / Trudy mezhdunarodnoj nauchno-prakticheskoj konferencii (14-15 nojabrja 2007 g., Moskva, Rossija). Obshhaja redakcija – V.N. Burkov, D.A. Novikov. – M.: IPU RAN, 2007. – S.186-190.
117. Fishbern P. Teorija poleznosti dlja prinjatija reshenij. – M.: Nauka, 1978. – 352 s.
118. Kiselev N.I. Jekspertno-statisticheskij metod opredelenija funkicii predpochtenija po rezul'tatam parnyh sravnenij ob#ektov // Algoritmicheskoe i programmnoe obespechenie prikladnogo statisticheskogo analiza. – M.: Nauka, 1980. – S.111-123.
119. Shahnov I.F. Nekotorye modeli kvalimetricheskogo analiza mnogofaktornyh ob#ektov s binarnymi faktorami // Zavodskaja laboratorija. Diagnostika materialov. 2005. T.71. №5. S.59-65.
120. Podinovskij V.V. Analiz reshenij pri mnozhestvennyh ocenках koeficientov vazhnosti kriteriev i verojatnostej znachenij neopredelennyh faktorov v celevoj funkicii // Avtomatika i telemekhanika. 2004. №11. S.141-159.
121. Podinovskij V.V. Kolichestvennaja vazhnost' kriteriev s nepreryvnoj shkaloj pervoj porjadkovoju metriki // Avtomatika i telemekhanika. 2005. №9. S.129-137.
122. Podinovskij V.V. Vvedenie v teoriju vazhnosti kriteriev v mnogokriterial'nyh zadachah prinjatija reshenij. – M.: Fizmatlit, 2007. – 64 s.
123. Gel'fand I.M., Alekseevskaja M.A., Guberman Sh.A. i dr. Prognozirovanie ishoda infarkta miokarda s pomoshh'ju programmy «Kora-3» // Kardiologija. 1977. T.17. №6. S.19-23.
124. Blekujell D., Girshik M. Teorija igr i statisticheskikh reshenij. - M.: IL, 1958. - 374 s.
125. Kini R.L., Rajfa H. Prinjatie reshenij pri mnogih kriterijah: predpochtenija i zameshhenija. – M.: Radio i svjaz', 1981. – 560 s.
126. Orlov A.I. Vosstanovlenie zavisimosti metodom naimen'shix kvadratov na osnove neparametricheskoj modeli s periodicheskoj sostavljajushhej // Nauchnyj zhurnal KubGAU [Elektronnyj resurs]. – Krasnodar: KubGAU, 2013. – №07(091). S.189–218. – IDA [article ID]: 0911307013. – Rezhim dostupa: <http://ej.kubagro.ru/2013/07/pdf/13.pdf>.
127. Lucenko E.V. Teoreticheskie osnovy i tehnologija adaptivnogo semanticheskogo analiza v podderzhke prinjatija reshenij (Na primere universal'noj avtomatizirovannoj sistemy raspoznavanija obrazov «JeJDOS-5.1») Monografija (nauchnoe izdanie). – Krasnodar: KJuI MVD RF, 1996. – 278 s.
128. Lucenko E.V., Simankov V.S. Adaptivnoe upravlenie slozhnymi sistemami na osnove teorii raspoznavanija obrazov. Monografija (nauchnoe izdanie). – Krasnodar: TU KubGTU, 1999. –318 s.
129. Orlov A.I., Tolcheev V.O. Ob ispol'zovanii neparametricheskikh statisticheskikh kriteriev dlja ocenki tochnosti metodov klassifikacii (obobshhajushhaja stat'ja) // Zavodskaja laboratorija. Diagnostika materialov. 2011. T.77. №3. S.58-66.

130. Orlov A.I. Ocenki plotnosti v prostranstvah proizvol'noj prirody // Statisticheskie metody ocenivaniya i proverki gipotez: mezhvuz. sb. nauch. tr. / Perm. gos. nac. issl. un-t. – Perm', 2013. – Vyp. 25. – S.21-33.
131. Orlov A.I. Prognosticheskaja sila kak pokazatel' kachestva algoritma diagnostiki. - V sb.: Statisticheskie metody ocenivaniya i proverki gipotez: mezhvuz. sb. nauch. tr. Vyp.23. – Perm': Perm. gos. nac. issl. un-t, 2011. – S.104-116.
132. Savchenko A.V. Metod napravlenogo perebora al'ternativ v zadachah raspoznavaniya obrazov // Zavodskaja laboratorija. Diagnostika materialov. 2011. T.77. №11. S.69-74.
133. Savchenko A.V. Kombinirovaniya diagnosticheskikh pravil na osnove informacionnogo rassoglasovanija Kul'baka-Lejblera // Zavodskaja laboratorija. Diagnostika materialov. 2012. T.78. S.69-73.
134. Savchenko A. V. Ob odnom sposobe povyshenija nadezhnosti reshenija zadachi raspoznavaniya obrazov. // Zavodskaja laboratorija. Diagnostika materialov. 2013. T.79. №1. S.72-77.
135. Gorbach A.N., Cejtin N.A. Klasterizacija spontannyh posledovatel'nostej // Zavodskaja laboratorija. Diagnostika materialov. 2009. T.75. №5. S.64-69.
136. Gorbach A.N., Cejtin N.A. Algoritm klasterizacii spontannyh posledovatel'nostej // Zavodskaja laboratorija. Diagnostika materialov. 2010. T.76. №1. S.69-70.
137. Borodkin A.A., Tolcheev V.O. Kompleksnaja procedura redukcii dlja uvelichenija bystrodejstvija neparametricheskikh metodov klassifikacii tekstovyh dokumentov // Zavodskaja laboratorija. Diagnostika materialov. 2011. T.77. №11. S.64-69.
138. Borodkin A.A., Tolcheev V.O. Razrabotka i issledovanie metodov vzveshivaniya blizhajshih sosedej (na primere klassifikacii bibliograficheskikh tekstovyh dokumentov) // Zavodskaja laboratorija. Diagnostika materialov. 2013. T.79. №7. S.70-74.
139. Orlov A.I. Ustojchivost' klastera – kriterij ego estestvennosti // Matematicheskie metody izuchenija geologicheskikh javlenij [Sb. st.] Mosk. ob-vo ispytatelej prirody, Mezhshekc. seminar po primeneniju matematiki v geologii; [Gl. red. A. L. Janshin]. - M.: Nauka/MOIP, 1990. – S. 54-60.
140. Orlov A.I. Ustojchivost' klassifikacii otnositel'no vybora metoda klaster-analiza // Zavodskaja laboratorija. Diagnostika materialov. 2013. T.79. №1. S.68-71.
141. Lucenko E.V. Cistemno-kognitivnyj analiz izobrazhenij (obobshhenie, abstragirovanie, klassifikacija i identifikacija) // Nauchnyj zhurnal KubGAU [Elektronnyj resurs]. – Krasnodar: KubGAU, 2009. – №02(46). – Shifr Informregistra: 0420900012\0017. – Rezhim dostupa: <http://ej.kubagro.ru/2009/02/pdf/10.pdf>
142. Makarkin A. A. Raspoznavanie ob#ektov po tupikovym testoram // Zavodskaja laboratorija. Diagnostika materialov. 2013. T.79. №3. S.63-71.
143. Badanov A.V., Borisova I.A., Djubanov V.V., Zagorujko N.G., Kutnenko O.A., Kuchkin A.V., Meshherjakov M.A., Milozvorov N.G. Sistema SPEKTRAN dlja intellektual'nogo analiza spektral'nyh dannyh // Zavodskaja laboratorija. Diagnostika materialov. 2009. T.75. №11. S.65-70.
144. Orlov A.I. Novaja hronologija kak osnova gosudarstvenno-patrioticheskogo mirovozzrenija // Nauchnyj jekspert. – 2013. – №3. – S.76-87. – Rezhim dostupa: http://problemanalysis.ru/text/Jornal3_2013.pdf.
145. Lucenko E.V., Trunев A.P. Fundamental'nye zakonomernosti raspoznavaniya social'nyh kategorij po astronomicheskim dannym na moment rozhdenija // Nauchnyj zhurnal KubGAU [Elektronnyj resurs]. – Krasnodar: KubGAU, 2008. – №10(44). – Shifr Informregistra: 0420800012\0138. – Rezhim dostupa: <http://ej.kubagro.ru/2008/10/pdf/01.pdf>

146. Orlov A.I. O razvitii matematicheskikh metodov teorii klassifikacii // Zavodskaja laboratorija. Diagnostika materialov. 2009. T.75. №7. S.51-63.
147. Novikov D.A., Orlov A.I. Matematicheskie metody klassifikacii // Zavodskaja laboratorija. Diagnostika materialov. 2012. T.78. №4. S.3-5.
148. Orlov A.I. O razvitii statistiki ob#ektov nechislovoj prirody / A.I. Orlov // Nauchnyj zhurnal KubGAU [Jelektronnyj resurs]. – Krasnodar: KubGAU, 2013. – №09(093). S. 273 – 309. – IDA [article ID]: 0931309019. – Rezhim dostupa: <http://ej.kubagro.ru/2013/09/pdf/19.pdf>.
149. Orlov A.I., Lucenko E.V. Sistemnaja nechetskaja interval'naja matematika (SNIM) – perspektivnoe napravlenie teoreticheskoy i vychislitel'noj matematiki // Nauchnyj zhurnal KubGAU [Jelektronnyj resurs]. – Krasnodar: KubGAU, 2013. – №07(091). S. 255 – 308. – IDA [article ID]: 0911307015. – Rezhim dostupa: <http://ej.kubagro.ru/2013/07/pdf/15.pdf>.
150. Lucenko E.V., Korzhakov V.E. Metod kognitivnoj klasterizacii ili klasterizacija na osnove znaniy (klasterizacija v sistemno-kognitivnom analize i intellektual'noj sisteme «Jejdos») // Nauchnyj zhurnal KubGAU [Jelektronnyj resurs]. – Krasnodar: KubGAU, 2011. – №07(071). S. 528 – 576. – Shifr Informregistra: 0421100012\0253, IDA [article ID]: 0711107040. – Rezhim dostupa: <http://ej.kubagro.ru/2011/07/pdf/40.pdf>.