

УДК 681.142/155

ВТОРОЕ ВСЕСОЮЗНОЕ СОВЕЩАНИЕ И ШКОЛА-СЕМИНАР
ПО МАШИННЫМ МЕТОДАМ ОБНАРУЖЕНИЯ ЗАКОНОМЕРНОСТЕЙ (МОЗ-2)

Н.Г.Загоруйко, Л.А.Растригин

С 26 февраля по 2 марта в г.Юрмале и г.Риге (ЛатвССР) проходили Второе Всесоюзное совещание и школа-семинар по проблеме "Машинные методы обнаружения закономерностей" (МОЗ-2). Они были организованы Латвийским республиканским правлением НТОРЭС им.А.С.Попова совместно с Институтом математики СО АН СССР. На совещании и школе-семинаре присутствовали представители 30 организаций из 18 городов, среди которых - Москва, Новосибирск, Киев, Минск, Рига, Свердловск, Тбилиси, Ереван, Одесса, Владивосток, Ташкент.

Обсуждались методологические, алгоритмические и прикладные аспекты проблемы обнаружения закономерностей. Работа проходила в два этапа: совещание с участием узкого круга специалистов и школа, рассчитанная на более широкий круг участников.

На совещании было прочитано 22 доклада по актуальным проблемам МОЗ. Во вступительном слове Л.А.Растригин (Рига) остановился на вопросе о месте проблематики МОЗ в общей структуре научного знания. В терминах размытых представлений положение тематики МОЗа было определено между идентификацией и прикладной гносеологией.

В докладе Н.Г.Загоруйко (Новосибирск) "Машинные методы обнаружения закономерностей. Состояние проблемы" был приведен обзор исследований по проблеме МОЗ. Закономерность определялась как обладающая определенными свойствами гипотеза $\langle OT \rangle$ о том, что если мы будем наблюдать объекты методом O , то увидим только то, что согласуется с утверждением T . Отмечались такие свойства гипотез, как потенциальная опровержимость, степень подтвержденности, простота формулировки и т.п. Рассматривались проблема усиления закономерностей и проблема их использования для эмпирического предсказания.

Д.А.Воронин (Новосибирск) в докладе "Некоторые попытки введения понятия о МОЗ и геологические применения" рассмотрел ряд определений закономерностей, их использование в распознавании образов, иллюстрируя изложение геологическими примерами.

В ряде докладов были рассмотрены различные методологические аспекты проблемы обнаружения закономерностей. Доклад Е.Е.Витяева, К.Ф.Самохвалова, Д.И.Свириденко (Новосибирск) "Проблемы теории измерений" содержал критический анализ современной репрезентативной теории измерений. Предлагалось рассматривать концепцию измерения в рамках новой аксиоматической теории измерений, обходящей отмеченные недостатки. Были введены основные понятия данной теории и в их терминах предложено решение некоторых проблем измерений.

В докладе Е.Е.Витяева (Новосибирск) "Закономерности в языках эмпирических систем" рассматривался подход, основанный на замене признаков, характеризующих объекты в задачах распознавания и эмпирического предсказания, на эмпирические системы этих признаков. Описан метод предсказания, использующий многоместные отношения из эмпирических систем. Показано представление некоторых законов физики посредством формул, записанных в терминах многоместных отношений. Отмечалось, что некоторые закономерности, использующие двуместные отношения, логически несводимы к закономерностям, использующим одноместные отношения.

К.Ф.Самохвалов (Новосибирск) в докладе "Роль теоретических терминов в аксиоматических теориях" провел сравнение двух методов представления эмпирических теорий: аксиоматического и алгоритмического. Показано, что некоторые пороки аксиоматических теорий, обычно связываемые с наличием в них теоретических терминов, сохраняются и при отсутствии таковых. От этих пороков свободен алгоритмический подход к представлению эмпирических теорий.

Доклад Б.П.Гаврилко (Новосибирск) "Семантический подход к теории эмпирического предсказания" был посвящен методологическим аспектам проблемы обнаружения эмпирических закономерностей. Предложены изменения представлений известного синтаксического подхода, использующего понятие теории эмпирического предсказания. Изменения затрагивают понятие эмпирического объекта и тесно связанное с ним понятие наблюдения. Показано, что семантический подход позволяет избежать трудностей, обнаруженных в этой области при синтаксическом подходе.

В докладе Г.П.Щедровицкого (Москва) "Естественное и искусственное в обнаружении научных законов" научные законы и закономерности рассматривались как особые формы знаний, функционирующих в

сложных структурах деятельности. Было показано, что решающую роль в оформлении знаний в качестве законов и закономерностей играют формы и способы использования этих знаний. Анализируя далее связь двух систем деятельности, в которых знание сначала порождается, а затем используется, докладчик показал, что закономерности и законы - это особые "технические" правила организации нашей деятельности и мышления, замещающие натуральные механизмы изменения измеряемых нами величин.

В докладе Г.С.Цветкова (Рига) "Опыт поиска закономерностей в проектировании управляющих систем" закон рассматривался как особого типа знание, интерпретированное на реальный объект. Приведен процесс выработки теоретико-структурного представления "мира автоматизированной деятельности", и дан его логический анализ. На материале практической разработки рассмотрены особенности эмпирического, теоретического и методологического подходов в обнаружении закономерностей, а также основные результаты в виде, в частности, функционально-структурного подхода и "языка описания", базирующегося на нем.

Л.Л.Леоненко (Одесса) в докладе "Об одном формальном методе системного представления объектов" изложил метод представления реальных объектов в виде систем, основанный на формализации имеющейся информации об этих объектах в языке тернарного описания.

Алгоритмические вопросы обнаружения закономерностей были темой наиболее представительной группы докладов. Г.С.Лбов (Новосибирск) в докладе "Логические функции в задачах эмпирического предсказания" представил формулировку подхода, использующего логические функции для решения различных задач, возникающих при обработке эмпирических таблиц, характерной особенностью которых является наличие большого числа признаков, измеренных в шкалах разных типов. При этом число объектов сравнимо с числом признаков, допускаются пропуски в таблицах. Рассмотрены преимущества представленного подхода для анализа указанного класса таблиц по сравнению с известными при решении задач распознавания образов, восстановления функции, автоматической группировки объектов, динамического прогнозирования.

Логические методы обнаружения закономерностей рассматривались в докладе А.Д.Закревского (Минск) "Логические модели исследуемого окружения и алгоритмы их построения" на примере задачи обнаружения закономерностей в "неслучайном" подмножестве ME булева пространства двоичных признаков. Предложено аппроксимировать

это подмножество силлогической модели - системой простых суждений, эквивалентной д.н.ф. второго порядка. Рассмотрены обобщения, соответствующие повышению порядка д.н.ф., и сформулированы комбинаторные задачи, связанные с оценкой достоверности моделей, конструируемые по ограниченным случайным выборкам из МЕ.

В докладе В.Е.Голендера (Рига) "Индуктивная логика и обнаружение закономерностей" отмечалось, что на базе миллевских методов индуктивной логики в сочетании с системами представления проблемно-ориентированного знания могут быть построены эффективные алгоритмы обнаружения закономерностей. Такой подход фактически интуитивно используется человеком при решении соответствующих задач, на нем базируются и многие известные алгоритмы. В качестве иллюстрации был рассмотрен ряд примеров решения геометрических и алгебраических психологических тестов, связанных с обнаружением закономерностей.

Б.Я.Ковалерчук (Ташкент) в докладе "Вопросы получения априорной информации и обнаружения закономерностей в многокритериальных задачах" в рамках утверждения, что единственным средством обоснования выбора одной из одинаково подтвержденных закономерностей является привлечение априорной информации, рассмотрел задачи разработки процедур: 1) диалога для получения априорной информации, 2) формального вывода из полученной информации следствий, необходимых для выбора закономерности. Предложен подход к решению указанных задач путем построения слабых шкал на базе слов естественного языка, формализации в виде формул неклассического исчисления $\dot{I}$ с императивом и заменой вывода в I на решение задачи оптимизации в RH .

Алгоритм обнаружения закономерностей, содержащихся в последовательности символов, был предложен М.И.Шлезингером (Киев) в докладе "Представление знаний с помощью двумерных грамматик". В основе алгоритма лежит аппарат двумерных грамматик и ранее разработанный автором алгоритм самообучения. Алгоритм опробован как на искусственных примерах, так и анализе буквенных последовательностей, представляющих газетный текст.

Л.А.Растрегин и Д.В.Самсонс (Рига) в сообщении "Метод многомерной линейной экстраполяции в задачах выявления динамических закономерностей и восстановления таблицы" был описан алгоритм построения модели динамического процесса, представленного выборкой наблюдений. Эта модель строится с использованием метода многомерной линейной экстраполяции и может быть использована для опреде -

ления поведения объекта в новых ситуациях. Преимуществом метода является простота учета большого числа входов объекта. Авторы описали также алгоритм определения недостающих элементов избыточной системы путем решения специальным образом построенной минимизационной задачи. Минимизируемый критерий образуется как суммарная невязка элементов таблицы и их экстраполированных значений, полученных методом многомерной линейной экстраполяции. Полученная задача имеет обычно многоэкстремальный характер и эффективно решается методами случайного поиска.

Метод многомерной линейной экстраполяции для синтеза решающих правил рассматривался в докладе М.И.Фабриканта (Рига) "Априорный синтез экстраполяционного решающего правила на основе классификации выборки".

Применению метода группового учета аргументов был посвящен доклад А.Г.Ивахненко и А.Н.Грошкова (Киев) "Обнаружение закономерностей физических объектов при помощи перебора по критерию минимального смещения".

В докладе Н.А.Неваля (Рига) "Об одном подходе к построению алгоритмов обнаружения скрытых закономерностей" был изложен подход, в котором под скрытыми закономерностями понимались: зависимость одной физической величины от другой, момент "разладки" процесса, последующее значение порядковой статистики. В основу подхода положена идея инвариантного "связывания" неизвестных параметров достаточными статистиками, что позволяет получить новые и важные для практического приложения результаты.

Группа из трех докладов была посвящена одному из видов индуктивного вывода - предельной идентификации функций. Р.В.Фрейвалд (Рига) в докладе "Вероятностная предельная идентификация общерекурсивных функций в различных вычислимых нумерациях" провел сравнение возможностей вероятностных и детерминированных стратегий восстановления номеров общерекурсивных функций в различных вычислимых нумерациях частично-рекурсивных функций по значениям этих функций. Метод получения естественных примеров предельно идентифицируемых классов функций был изложен в сообщении Б.Е.Кинбера (Рига) "О некоторых проблемах предельной идентификации функций", построен также предельно идентифицируемый класс функций, не содержащий эффективно перечислимых подклассов. А.В.Анджан (Рига) в сообщении "Предельный синтез и прогнозирование частично-рекурсивных функций" изложил различные возможности обобщения понятий предельного синтеза и прогнозирования классов общерекурсивных функций на

случай классов частично-рекурсивных функций. Построена полная картина включения-невключения для семейств прогнозируемых (синтезируемых) функций.

В ряде докладов были рассмотрены различные аспекты проблемы распознавания образов. Н.Г.Загоруйко в докладе "Проблемы таксономии" сформулировал эту задачу, как задачу минимизации потерь, связанных с определенным видом заменами одних объектов другими, входящими в один и тот же таксон. Был приведен обзор методов таксономии с этих позиций.

Ш.Д.Раудис (Вильнюс) в докладе "Соотношение объема выборки, числа признаков и сложности правила с ошибками классификации" привел количественную и качественную связь между ошибкой классификации, объемом обучающей выборки, числом признаков и сложностью решающего правила для класса решающих правил, основанных на нормальном распределении признаков. Показано, что оптимизация построения решающего алгоритма при конечном объеме выборки должна идти по пути одновременного выбора решающего правила и системы признаков.

Подход к распознаванию объектов-графов, основанный на отборе признаков, представляющих собой подграфы, общие для совокупности объектов одного класса, был представлен в докладе Л.С.Гитлиной, В.Е.Голендера и А.Б.Розенблита (Рига) "Поиск закономерностей в классах объектов, представленных графами". Предложен алгоритм отбора признаков-подграфов, и рассмотрено его применение для классификации структур химических соединений по их активности.

В сообщении В.В.Джанкиери (Тбилиси) и Л.А.Растрюгина (Рига) "Оптимизация решающих правил" описаны структурно-параметрические методы оптимизации решающих правил распознавания образов. Один из подходов связан с организацией коллектива решающих правил, а другой - с образованием гибридных правил. В обоих случаях оптимизация по параметрам коллектива и свертки позволяет значительно повысить эффективность решения. При отсутствии параметров в решающих правилах предлагается случайно варьировать обучающей выборкой, что обеспечивает и коллектив и гибриды достаточным числом правил с требуемым разнообразием.

А.Б.Глаз и М.И.Коган (Рига) в докладе "Построение решающего правила на недостаточной обучающей выборке" рассмотрели понятие недостаточной выборки, для которой условие минимума эмпирического риска не гарантирует оптимальности построенного в процессе обучения решающего правила. Для повышения экстраполяционных возможностей введена дополнительная процедура голосования по множеству

отдельных решающих правил, построенных из условия минимума эмпирического риска. На примере конкретных задач показана эффективность такого подхода.

В докладе Р.С.Тольдмана (Владивосток) "Модели нечеткой логики в методах поиска закономерностей" были введены понятия нечетких тестов, тупиковых нечетких тестов и рассмотрены методы их построения с помощью операций нечеткой логики.

В докладе В.М.Величко и В.Д.Гусева (Новосибирск) "Использование ассоциативного кодирования для вычисления мер близости между отдельными объектами и их совокупностями" были рассмотрены процедуры типа "хеширования" (расстановки, перемешивания), которые являются удобным аппаратом для вычисления мер близости между отдельными объектами и совокупностями объектов. Иллюстрацией служили задачи поиска "ближайших соседей" (трудоемкость T не зависит от числа объектов в совокупности) и вычисления меры близости между двумя перестановками N объектов ($T=O(N)$).

В.К.Финном и М.И.Забейало (Москва) в докладе "О машинной программе классификации трехзначных логик некоторого класса" была изложена реализованная на ЭВМ программа автоматической классификации, с помощью которой удалось доказать некоторую теорему из алгебры логики.

Доклад А.И.Половинкина (Йошкар-Ола) "Проблема машинного поиска закономерностей, использования физических закономерностей (типа физических эффектов) в технических системах" был посвящен анализу закономерностей, используемых в рассмотренной предметной области. Предложена следующая иерархия закономерностей: а) частные закономерности в виде отдельных физических принципов действия, которые широко применяются в технике; б) типовые физические структуры, объединяющие многие принципы действия; в) всеобщие закономерности (законы) построения и функционирования технических систем. Было высказано предположение о целесообразности использования алгоритмов и программ обнаружения закономерностей при формулировке указанных закономерностей.

На состоявшейся после совещания школе-семинаре было заслушано 7 лекций, 5 докладов и 24 сообщения.

В лекции Н.Г.Загоруйко (Новосибирск) был дан обзор состояния проблемы МОЗ и описан пакет прикладных программ для обработки таблиц экспериментальных данных ОТЭС. Функциональные блоки этого

пакета решают определенные задачи обнаружения эмпирических закономерностей.

Г.П.Щедровицкий (Москва) в лекции "Естественное и искусственное в выявлении законов и закономерностей" на ряде исторических примеров проиллюстрировал общие положения, изложенные в докладе на совещании.

Прикладной теории измерений была посвящена лекция Б.П.Гаврилко (Новосибирск).

Различные подходы к проблеме заполнения пропусков в таблицах данных были изложены в лекции Ю.П.Дробышева и С.В.Макарова (Новосибирск) "К проблеме обработки неполных данных. (Обзор и состояние)". Было проведено сравнение методов регрессии и главных компонент с некоторыми эвристическими приемами. Указаны новые интерпретации регрессионного подхода.

В лекции Ю.П.Пономарева и Л.А.Растригина (Рига) "Многомерная линейная экстраполяция как метод обнаружения закономерностей" был описан метод, позволяющий по обучающей выборке строить непараметрическую модель. Преимуществом этого подхода является возможность обнаруживать закономерности с большим числом входов при малом числе наблюдений. Описаны практические примеры выявления закономерностей из области проектирования, оперативного управления и идентификации динамических систем.

Проблемы МОЗ в техническом конструировании были проанализированы в лекции А.И.Половинкина и Е.А.Смиренского. Вопросам распознавания образов были посвящены лекция Г.С.Лбова (Новосибирск) "Логические решающие функции" и доклад А.Б.Глаза (Рига) "Синтез оптимального решающего правила на достаточной обучающей выборке". Конкретным системам для решения задач распознавания образов и обнаружения закономерностей были посвящены доклады Ш.Д.Раудиса (Вильнюс) "Система СОРРА", В.П.Гладуна и Н.Д.Ващенко (Киев) "АНАЛИЗАТОР - система выделения закономерностей", А.Б.Розенблита и В.Е.Голендера (Рига) "Системы программ для поиска закономерностей химическая структура - биологическая активность".

В докладе Э.М.Погосяна, Н.А.Аджабяна, Л.О.Диндоляна, Ю.Г.Арутюняна, М.А.Амбарцумяна (Ереван) "Пакет прикладных программ для исследования методов автоматического усиления стратегий управления" на примере шахматной игры исследованы возможности доказательства либо опровержения гипотез, связанных с автоматическим усилением стратегий управления.

Методологические вопросы обнаружения закономерностей рассматривались в сообщениях М.Б.Гнедовского "О фиксации законов системного типа", Л.Кузнецовой и С.Поливановой (Москва) "Анализ проблем обработки экспериментальных данных", Савусина (Одесса) "Системно-параметрический критерий близости процессов познавательной деятельности по их сложности".

Ряд сообщений: А.В.Глаз, М.И.Коган (Рига) "Учет априорной информации в задачах распознавания образов", Э.М.Кадырбаева, Г.П.Кнебелис, Я.И.Приживойт (Фрунзе) "Диалоговая система алгоритмов автоматической классификации", М.А.Веревичева, В.И.Мелешко (Харьков) "О сопоставлении трех методов распознавания образов", В.В.Джандиери (Тбилиси) "Библиотека программ распознавания" был посвящен различным проблемам распознавания образов.

В сообщениях В.В.Разникова (Черноголовка) "Об одном методе прогнозирования свойств органических соединений", Е.М.Савицкого, Н.Н.Кисилевой (Москва) "Поиск закономерностей образования химических соединений в двойных и тройных неорганических системах с помощью обучения ЭВМ", Д.А.Неймарка, Л.Г.Теклиной (Горький) "О выделении закономерностей заданной формы", Р.С.Гольдмана, В.В.Николаева, Ю.С.Москаленко (Владивосток) "Прогноз сроков подхода рыбопромысловых скоплений сайры с использованием методов обнаружения закономерностей", А.Ф.Аба (Рига) "Регрессионные модели показателей функционирования режимов ОЭС на основе полиномов Чебышева в форме Нейчинова", А.Н.Дурново (Горький) "Методы обнаружения скрытых закономерностей процессов функционирования сложных объектов", В.Д.Гусева, Д.Г.Косарева, М.К.Тимофеевой, Т.Н.Титковой (Новосибирск) "Об автоматическом обнаружении устойчивых единиц текста" были рассмотрены алгоритмы и результаты решения конкретных задач.

Различные алгоритмы обнаружения закономерностей рассматривались в сообщениях А.Н.Борисова, О.А.Крумберга (Рига) "Алгоритмы восстановления субъективных закономерностей", Н.И.Калядина, А.И.Муренова (Ижевск) "Обнаружение классификационных закономерностей методом альтернативных решающих правил", В.А.Агладзе, Д.П.Пономарева (Рига) "Рекурсивная форма метода многомерной линейной экстраполяции", В.И.Котикова, А.Е.Буторина (Новосибирск) "Синтез кусочно-линейных моделей и их структурная адаптация", В.Н.Пухова, Д.П.Дробышева (Новосибирск) "Преобразование таблиц в анализе данных", Д.В.Каширского, Д.И.Астахова "Методика комплексного машинного анализа экспериментальных данных", В.В.Качала (Пушкино) "Новые алгоритмы идентификация - развитие метода группового учета аргу-

ментов", С.С.Фомина (Калининград) "Один из методов обнаружения закономерностей в многомерном случайном процессе", Г.А.Андриксона (Рига) "О применении эргодической гипотезы при использовании прикладных методов теории случайных процессов".