

ПРОГРАММНАЯ СИСТЕМА ПОИСКА ОПТИМАЛЬНЫХ РЕГРЕССИЙ

С.Г. Валеев, Г.Р. Кадырова, А.А. Турченко

ГОУ ВПО «Ульяновский государственный технический университет», г. Ульяновск

Рецензент Г.М. Куликов

Ключевые слова и фразы: методы структурной идентификации; программный продукт; прогнозирование; регрессионное моделирование.

Аннотация: Описывается программный пакет «Система поиска оптимальных регрессий», позволяющий осуществлять высокоточное статистическое (регрессионное) моделирование процессов или явлений с последующим использованием моделей для прогноза выходных характеристик.

Введение. Развитие техники, получение высокоточных наблюдений искусственных и естественных небесных тел сделало чрезвычайно актуальной задачу создания пакетов прикладных программ (ППП), с помощью которых можно было бы проводить прецизионную математическую обработку результатов наблюдений.

Многоцелевой характер ППП, то есть нацеленность их на решение ряда задач по крупным разделам прикладной статистики, является без сомнения достоинством автоматизированных систем в условиях охвата многочисленных практических проблем отраслей наук. Однако, с другой стороны, это свойство ограничивает применение ППП как в программном смысле (по объему памяти и количеству неиспользуемых модулей), так и в математическом (скудное «меню», например, по методам оценивания). Для прецизионной астрономии из задач, рассматриваемых в прикладной статистике, наиболее характерны задачи оценивания методом наименьших квадратов (МНК) и структурной идентификации при регрессионном анализе (РА). Следовательно, основной проблемой в этой области является проблема создания специализированного ППП по РА МНК,

Валеев С.Г. – доктор физико-математических наук, профессор, заведующий кафедрой «Прикладная математика и информатика» УлГТУ; Кадырова Г.Р. – кандидат технических наук, доцент кафедры «Прикладная математика и информатика» УлГТУ; Турченко А.А. – аспирант кафедры «Прикладная математика и информатика» УлГТУ, г. Ульяновск.

функциональными модулями которого могли бы быть в основном оригинальные подпрограммы.

Программный пакет СПОР. Разработанный программный пакет «Система поиска оптимальных регрессий» (СПОР) является специализированной системой, реализующей стратегию статистического (регрессионного) моделирования [1]. Основное назначение системы – получение регрессионных моделей процессов или явлений с последующим их использованием для прогноза выходных характеристик (откликов) и реализации некоторых функций управления в интерактивном (дисплейном) и пакетном режимах работы. Необходимость наличия подобной системы порождается большими затруднениями при выполнении подобных работ, требующих как многовариантности расчетов, так и применения различных методов оценивания параметров и структурной идентификации, а также анализа остатков при выбранном сценарии проверки соблюдения предположений МНК и последующей адаптации при их нарушении.

Область применения. Пакет СПОР может использоваться для решения:

- задач метода наименьших квадратов (задач восстановления зависимостей по избыточным косвенным наблюдениям) и регрессионного анализа в любых областях (геодезии, фотограмметрии, астрометрии, небесной механике, технике, экономике и т.д.);
- различных технологических задач «черного ящика», требующих восстановления эмпирической зависимости между выходным параметром процесса и набором входных;
- задач трансформации координат при построении геодезических, геодинамических и других специализированных сетей.

Функциональное наполнение пакета СПОР. Пакет обеспечивает выполнение следующих функций: формирование постулируемой модели на основе исходного или построенного полинома; структурно-параметрическую идентификацию: множественная линейная регрессия, гребневая регрессия, полный перебор структур, неполный перебор структур, перебор нормальных систем, пошаговая регрессия с включением-исключением, случайный поиск с адаптацией, случайный поиск с возвратом; поиск множества оптимальных моделей по заданному критерию с заданным уровнем значимости; построение диаграммы рассеяния; построение и анализ графиков остатков.

Дополнительный инструментарий пакета: встроенный редактор, построитель сравнительных таблиц, тонкая настройка работы пакета, модуль работы со сценариями, инструмент построения диаграммы рассеяния.

В рамках пакета СПОР реализована возможность быстрого формирования сценария будущего диалога с автоматическим встраиванием диалога в проблемно-ориентированную область пакета. Для пользователя необходимо лишь создать файл с расширением *.sch согласно правилам, изложенным во встроенном справочнике. Пакет СПОР имеет несколько сценариев, которые можно менять. Эти файлы входят в стандартную комплектацию СПОР и находятся в директории Scheme.

Сценарий может содержать как основные методы и инструменты СПОР, так и дополнительные возможности, перечисленные в справочной

системе. Таким образом, реализованная в рамках СПОР возможность построения диалога конкретного сценария позволяет достаточно легко формировать всевозможные сценарии обработки данных.

Конкурентоспособность СПОР. Разработанная программная система конкурентоспособна по отношению к таким статистическим пакетам, как Statistica или Statgraphics при решении задач РА-МНК (при построении и оценивании регрессионных моделей) и обеспечивает получение оптимальной модели при использовании различных методов структурной идентификации не только по классическим (внутренним) мерам качества, но и внешним мерам и мерам скользящего экзамена, что позволяет осуществлять оценку внешней адекватности модели (точности прогноза).

Математическое обеспечение СПОР содержит новые подходы и методы: регрессионное моделирование (**РМ**-подход), методы оценивания по внешним мерам и мерам скользящего экзамена, методы структурной идентификации, такие как полный перебор, неполный перебор переопределенных и нормальных систем, многокритериальный метод пошаговой регрессии с включением-исключением, случайный поиск с адаптацией и случайный поиск с возвратом.

Версии СПОР. Первая версия программного пакета была реализована на языке Object Pascal с использованием библиотек Turbo Vision и функционировала под управлением MS DOS [2].

Версия СПОР 2.0 разработана на платформе Windows (Win9x, NT, XP), что позволяет решить проблему совместимости с различными операционными системами линейки Windows [3]. Программа оперирует всей доступной памятью компьютера под управлением операционной системы.

В новую версию СПОР внедрены внешние процедуры предыдущей версии: процедура построения диаграммы рассеяния и процедура построения и анализа графика остатков. Дополнительно встроен текстовый редактор с возможностью просмотра текста в кодировках WIN-1251 и OEM/DOS-866, что упростило работу с данными.

Переход на другую операционную систему повлек за собой кардинальные изменения в интерфейсе программы, основанные на передовых технологиях: многозадачности, Drag & Drop и др. Интерфейс программы является дружелюбным – интуитивно понятным.

В пакете предусмотрены три уровня защиты: контрольное измерение размера файла, применение упаковщиков файлов приложений (ASPack 2.12), применение лицензионных ключей.

При запуске программы сравнивается размер запускаемого файла с контрольным размером файла, устанавливаемым при компиляции проекта. Что исключает возможность простого изменения кода программы после компиляции. Для включения полных функциональных возможностей программы необходим лицензионный ключ. Лицензионный ключ включает в себя – имя пользователя лицензионного ключа, версию программы, для которой рассчитан данный лицензионный ключ, даты начала и конца активации лицензионного ключа.

Развитие СПОР. Одним из направлений развития системы стала реализация метода структурной идентификация – полного перебора для параллельных вычислений, предусматривающая:

– разработку соответствующей подсистемы на основе библиотеки MPI в реализации MPICH для выполнения вычислений на кластере (создание серверной и клиентской части системы, модуля балансировки нагрузки, модуля потоковой работы с файлами данных, интерфейсной части системы, скриптов запуска системы);

– адаптацию метода полного перебора структур для режима параллельных вычислений (модификация метода Хаусхолдера, учитывающая спецификацию библиотеки MPI; преобразование вида модели для ускорения обмена между элементами кластера и управляющим процессом).

Разработанная подсистема позволяет производить поиск оптимальной модели по заданному критерию методом полного перебора структур с использованием кластера (многопроцессорных систем).

Использованный при отладке кластер состоял из восьми рабочих станций и одного сервера. Для запуска использовалось 20 процессоров [1 сервер (4 процессора) + 8 рабочих станций (16 процессоров)]; управляющий процесс запускался на одной из рабочих станций; приоритетность задачи была среднего уровня, при этом все компьютеры кластера могли использоваться для обычной работы.

Анализ полученных результатов показал, что использование алгоритма с распределением вычислений позволяет уменьшить время вычислений примерно в n раз, где n – число элементов кластерной сети, по сравнению с выполнением аналогичного вычисления на одном рабочем компьютере с такой же конфигурацией.

Программа реализована на основе интерфейса взаимодействия MPI в реализации MPICH, что дает ей возможность легко масштабироваться. Функциональная часть построена с использованием языка Free Pascal, а интерфейсная (взаимодействие с пользователем) – в среде разработки Turbo Delphi.

Заключение. В перспективе планируется расширение функционального наполнения пакета, активное использование распределенных технологий, повышающих эффективность вычислений, интеллектуализация пакета, решение проблем интегрирования СПОР и информационных систем.

Список литературы

1. Валеев, С.Г. Регрессионное моделирование при обработке наблюдений / С.Г. Валеев. – М. : Наука, 1991. – 272 с. (Валеев, С.Г. Регрессионное моделирование при обработке данных / С.Г. Валеев. – 2-е изд., доп. и перераб. – Казань : ФЭН, 2001. – 296 с.)

2. Валеев, Г.Р. Система поиска оптимальных регрессий / Г.Р. Валеев, Г.Р. Кадырова. – Казань : ФЭН, 2003. – 160 с.

3. Валеев, Г.Р. Регрессионные модели на основе новой версии пакета СПОР / Г.Р. Валеев, Г.Р. Кадырова, А.А. Турченко // Тр. междун. конф. «Континуальные алгебраические логики, исчисления и нейроинформатика в науке и технике» ; Математические методы и модели в прикладных задачах науки и техники. – Т. 4. – Ульяновск : Изд-во Ул. гос. техн. ун-та. – 2006. – С. 77–79.

Software for Optimal Regressions Search

S.G. Valeev, G.R. Kadirova, A.A. Turchenko

Ulyanovsk State Technical University, Ulyanovsk

Key words and phrases: structural identification methods; programming product; forecasting; regression modeling.

Abstract: The paper describes the software package «The System of Optimal Regressions Search» which allows making a statistical precision (regression) modeling of processes or events with its subsequent using for forecasting of output characteristics.

© С.Г. Валеев, Г.Р. Кадырова, А.А. Турченко, 2008