

УДК 004.4, 004.8, 004.89

ПРОГРАММНЫЕ СИСТЕМЫ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ И ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ

В.Б. Звонков (*vladimirzvonkov802@yandex.ru*)
Сибирский государственный аэрокосмический
университет, Красноярск, Россия

В работе описывается функционал и назначение программных систем интеллектуального анализа данных и поддержки принятия решений, разработанных автором данной статьи. Описываются различные классы алгоритмов, включенных в программные системы. Имеется пошаговое руководство пользователя.

Введение

В современную эпоху информационных технологий, развития средств вычислительной техники и коммуникаций, увеличения объемов хранимой информации все острее назревает проблема обработки, систематизации и извлечения полезных знаний из больших объемов необработанной информации. Данное научное направление, называемое Data Mining, получило дальнейшее развитие и приложения в концепциях разработки средств искусственного интеллекта: моделирование разумного и социального поведения, моделирование когнитивных процессов, моделирование кооперации интеллектуальных агентов, решающих совместно в незнакомой среде сложную задачу, моделирование процессов человеческого мозга, зрения и т.д. С каждым годом все интенсивнее развивается рынок программных систем интеллектуального анализа данных и поддержки принятия решений, которые позволяют конечному пользователю применять математический аппарат для решения широкого спектра задач, не изучая смысл и суть работы математических методов. Распознавание образов и жестов, распознавание слитной речи, моделирование деятельности колоний муравьев и пчел, направленной на достижение единой цели – вот только некоторые примеры задач, для которых уже необходимо иметь в арсенале методы решения как ученому, так и аналитику, инженеру и просто современному человеку 21-го века, т.к. задачи анализа различной информации и принятия решений возникают почти каждый день. В этой связи, автору данной статьи было интересно разработать и реализовать различные методы Data Mining,

апробировать их при решении разнотипных задач реального мира, автоматизировать процесс их применения, сравнить полученные результаты и выработать рекомендации для конечных пользователей. Проведено сравнение авторских программных систем с уже существующими коммерческими аналогами на рынке. Некоторые исследования выполнены в [Звонков В.Б., 2011], [Звонков В.Б., 2012].

1. Общее описание алгоритмов, программных систем, аппаратного и программного обеспечения

Автором настоящей работы были разработаны модификации методов и реализованы в виде трех программных систем (Intelligence Technologies – Self-Adapting Genetic Algorithm (IT-SAGA) - № государственной регистрации программы для ЭВМ в Роспатенте 2011611120 от 3.02.2011, система автоматизированного нейросетевого и эволюционного моделирования, оптимизации, прогнозирования), являющихся как СППР, так и экспертными системами. Обозначенные программные системы включают в себя следующий алгоритмический инструментарий: стандартные и авторские генетические алгоритмы (ГА) для решения задач безусловной однокритериальной оптимизации [Звонков В.Б., 2011]; 5 методов учета ограничений [Michalewicz, Z. et al., 1996] для задач условной однокритериальной оптимизации (штрафные функции и специализированные операторы); стандартные и авторские ГА для решения задач безусловной и условной многокритериальной оптимизации [Звонков В.Б., 2011]; метод SPEA [Zitzler E. et al, 1999]; метод полного перебора; оригинальные процедуры локального поиска; многослойные перцептроны Ф. Розенблатта в общем виде с настраиваемыми параметрами; оптимизационные методы параметров искусственных нейронных сетей (ИНС); комитеты (арифметическое среднее и взвешенное арифметическое среднее) ИНС [Звонков В.Б., 2012]; ручной и автоматический выбор структур ИНС, активационных функций, пороговых значений с помощью модифицированных ГА; автоматизированное тестирование ГА и ИНС на множестве настроек и задач; вывод прогресса работы алгоритмов и автоматизированного тестирования, численных значений критериев эффективности, графиков изменения численных критериев эффективности, графиков сходимости для ГА, допустимых областей, постановок задач, множества и фронта Парето; вывод процесса обучения ИНС и процесса эволюции решений для ГА; графический пользовательский интерфейс; справочная система; файловые потоки ввода-вывода информации; вспомогательные функции для работы основного алгоритмического ядра вышеуказанных программных систем [Звонков В.Б., 2011], [Звонков В.Б., 2012]. Разработанные системы обеспечивают высокое быстродействие (т.к.

разрабатываемые алгоритмы многократно модифицировались вручную, а также проводилась оптимизация кода компилятором), высокую стабильность (проводилась отладка и статический анализ кода с помощью CppCheck) и эффективность работы, что было подтверждено продолжительными тестами (от 2-х до 4-х месяцев функционирования ПК без выключения каждое, на протяжении нескольких лет) в разных семействах операционных систем (Windows, Linux) и на 3-х разных конфигурациях ЭВМ (ноутбук Acer Extensa 5620, стационарный ПК с ЦП Intel Core i7-920 (Bloomfield) 2.66ГГц, 3.6 ГГц/ Cooler Master GeminII / Gigabyte GA-EX58-UD3R (rev. 1.0)/ БП Chieftec [APS-500S] 500W, ATX v2.2, EPS12V, Active PFC/ Radeon HD 4550 / ОЗУ 2 * DIMM DDR3 2048MB PC10666 1333MHz Kingston/ ЖД 3.5" Hitachi 160 Гб, стационарный ПК с ЦП AMD Athlon-64 2800+ 1.8 ГГц, 2.2 ГГц/ Radeon 9600 Pro / ОЗУ DDR 512 Мб / ЖД 3.5" Hitachi 160 Гб). Комплектующие в стандартной конфигурации (ноутбук) и в моих конфигурациях совместимы и стабильно работают в комплексе, что было подтверждено с помощью синтетических тестов и утилит (Memtest, Everest, 3DMark, PC Mark, Linpack, Performance Test, SiSoftware Sandra, SpeedFan, System Stability Tester, Video Memory Stress Test и др.) и математических расчетов. А также коллективом авторов и автором данной работы была разработана программная система «Автоматизированная система решения сложных задач глобальной оптимизации многоагентными стохастическими алгоритма (GOLEM-SA)», № государственной регистрации программы для ЭВМ в Роспатенте 2011611158 от 4.02.2011. Автором статьи также были реализованы метод Хука-Дживиса и PSO.

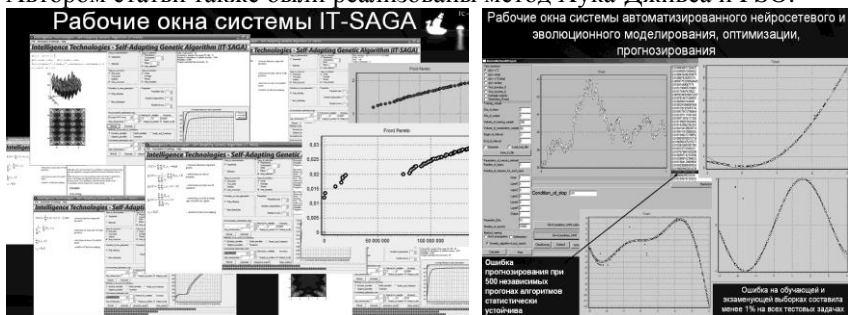


Рис. 1. Некоторые рабочие окна программных систем автора данной работы

2. Руководство пользователя программной системы IT-SAGA

После запуска программной системы IT-SAGA пользователю предлагается выбрать между двумя классами задач: однокритериальная и многокритериальная оптимизация.

Рассмотрим элементы рабочего окна программной системы (рис. 2).



Рис. 2 Основные элементы окна, ориентированного на решение задач класса однокритериальной оптимизации

2.1 Безусловная и условная однокритериальная оптимизация

2.1.1 Однокритериальная безусловная оптимизация.

В случае выбора тестовой задачи безусловной однокритериальной оптимизации в левой части окна отображается график целевой функции, аналитический вид и информация об экстремуме задачи, который определялся первоначально с помощью прикладного пакета MathCAD 14, а затем с помощью генетических алгоритмов.

2.1.2 Однокритериальная условная оптимизация.

В случае выбора тестовой задачи условной однокритериальной оптимизации в левой части окна отображается график целевой функции, аналитический вид, допустимая область и информация об экстремуме задачи, который определялся сначала с помощью MathCAD 14, а затем с помощью генетических алгоритмов. В случае выбора реальной задачи условной однокритериальной оптимизации в левой части окна отображается аналитический вид целевой функции и ограничений. По умолчанию, выбран метод учета ограничений «Динамические штрафы»

для задач условной оптимизации. В случае необходимости пользователь может выбрать другой метод: адаптивные штрафы, «смертельные» штрафы, «лечение», «лечение» + «смертельные» штрафы. Параметры всех методов учета ограничений для каждой задачи установлены оптимальными по умолчанию (получены при помощи многократных прогонов алгоритмов на данных задачах с перебором всех возможных комбинаций алгоритмических схем и обработкой результатов численных экспериментов по непараметрическим критериям в пакете Statistica).

2.1.3 Выбор параметров задачи.

Параметры задач безусловной и условной однокритериальной оптимизации, такие как область определения для вещественных переменных, шаг дискретизации для каждой вещественной переменной, длина хромосомы определены для каждой задачи по умолчанию. В случае необходимости пользователь может ввести другие числовые значения в соответствующие окна. При этом, длина хромосом при кодировании в бинарную строку рассчитывается автоматически.

2.1.4 Выбор параметров эволюционного алгоритма.

Если пользователь обладает необходимой теоретической информацией и достаточным опытом в области эволюционной оптимизации, априорной информацией о решаемой задаче, то он (она) может выбрать настройки эволюционного алгоритма вручную. Если пользователь такой информацией не обладает, то он (она) может обратиться к справочной системе программы и получить элементарные сведения об эволюционной оптимизации, а также выбрать автоматическую настройку основных параметров (флаги для автоматической настройки селекции, мутации и скрещивания установлены по умолчанию, существует возможность отключения). Рекомендуется выбрать размер популяции, количество поколений и количество прогонов алгоритма. Рекомендуемые начальные значения: 100 индивидов, 100 поколений, 500 статистически независимых прогонов.

2.1.5 Запуск решения задач (и).

Ручной режим работы программы. В рабочем окне необходимо выбрать из списка задач безусловную или условную оптимизацию (реальные и тестовые задачи). В данном режиме пользователь самостоятельно выбирает тип решаемой задачи, её параметры, при необходимости метод учета ограничений, настройки генетического алгоритма или автоматическую настройку. После выбора всех параметров необходимо нажать кнопку «**Manual**», соответствующую решаемому классу задач (безусловная или условная оптимизация), процесс работы на выбранной алгоритмической схеме отображает строка прогресса в нижней части окна. По окончании работы вся необходимая информация будет

отображена в окне результатов и (или) в выходном файле (после выполнения команды **File->Save**).

Автоматический режим работы программы. В данном режиме пользователь самостоятельно выбирает тип решаемой задачи (безусловная или условная однокритериальная оптимизация) установлением номера первой и последней решаемой задачи в списке выбранного класса задач, автоматически настраиваемые параметры генетического алгоритма (установлением или снятием соответствующих флагов автоматической настройки селекции, мутации, скрещивания). Для вывода дополнительной информации на экран необходимо установить флаг «**Output_to_window**», соответствующий решаемому классу задач, для вывода статистической информации в файл необходимо установить флаг «**Output_to_file**» (установлен по умолчанию), соответствующий решаемому классу задач. Рекомендуется выбрать размер популяции, количество поколений и количество прогонов алгоритма. После выбора всех параметров необходимо нажать кнопку «**Automatic**», соответствующую решаемому классу задач (безусловная или условная оптимизация), процесс работы отображают 3 строки прогресса (прогресс на текущей настройке, прогресс на множестве настроек в рамках текущей решаемой задачи, прогресс на множестве всех настроек на всех задачах выбранного списка) в нижней части окна. По окончании работы вся необходимая информация будет отображена в окне результатов и (или) в выходном файле.

2.2 Безусловная и условная многокритериальная оптимизация

Рассмотрим элементы рабочего окна программной системы (рис. 3). Алгоритм настройки и использования программной системы IT-SAGA для решения задач многокритериальной оптимизации аналогичен подробно описанной выше процедуре настройке и использования для решения задач однокритериальной оптимизации, и в силу ограниченности объема статьи не приводится. Отличие при решении задач данного класса состоит в том, что пользователю не нужно выбирать метод учета ограничений, для решения задач условной многокритериальной оптимизации достаточно включить соответствующим флагом алгоритм локального поиска. Решение задачи в этом случае осуществляется адаптивной гибридной стохастической процедурой с оптимальными параметрами по умолчанию.

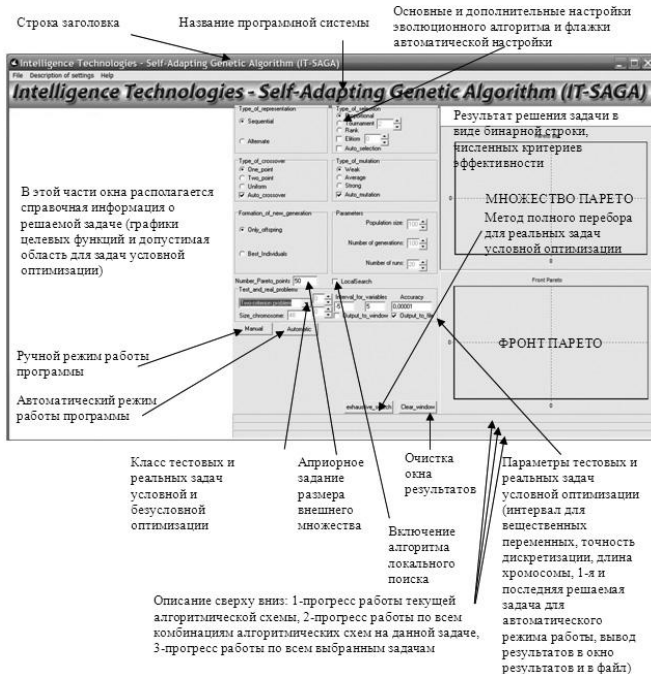


Рис. 3 Основные элементы окна, ориентированного на решение задач класса многокритериальной оптимизации

3. Руководство пользователя для программной системы эволюционного и нейросетевого моделирования, оптимизации, прогнозирования

Рабочее окно программной системы приведено на рис. 4. После запуска программной системы пользователю аналогично предлагается настройка параметров или выбор параметров по умолчанию, полученных автором данной работы. В окне программной системы пользователю доступны все необходимые настройки.

3.1. Выбор типа решаемой задачи

Тестовые функции (выборки генерируются или загружаются из файла) или реальные данные (выборки загружаются из файла).

3.2. Настройка параметров обучения и экзамена ИНС

Настраиваемые параметры: размерность входного вектора параметров (кол-во нейронов на 0-м слое); размерность выходного вектора (кол-во нейронов на последнем слое); кол-во точек в обучающей выборке (число примеров, на которых ИНС будут учиться решать поставленную задачу,

т.е. обучение с учителем); кол-во точек в экзаменуемой выборке (число примеров, на которых будет тестироваться качество функционирования обученных ИНС); границы области в вещественном пространстве для автоматической генерации выборки тестовых данных случайным образом; флаг «**Generate**» - генерация новой выборки для тестовых задач; флаг «**Load from file**» - загрузка существующей выборки из файла; кнопка «**Save to file**» - сохранение в файл текущей выборки, находящейся в памяти ЭВМ. Флаги «**Generate**» и «**Load from file**» начинают выполняться после запуска процесса обучения ИНС.



Рис. 4 Рабочее окно программной системы эволюционного и нейросетевого моделирования, оптимизации, прогнозирования после решения задачи прогнозирования курса валютной пары евро/рубль

3.3 Настройка вручную структуры ИНС

Настраиваемые параметры: кол-во скрытых слоев ИНС (кол-во внутренних слоев в ИНС между входным и выходным слоями); кол-во нейронов на каждом слое, находящемся между входным и выходным слоями; кол-во и тип активационных функций, пороговых значений. Некоторые параметры ИНС заданы по умолчанию в силу высокой трудоемкости настройки конечным пользователем.

3.4. Настройка оптимизационных методов обучения ИНС

3.4.1 Выбор условия останова процесса обучения ИНС.

Для алгоритма обратного распространения ошибки – максимальное количество эпох, или удовлетворение требуемой относительной ошибки (реализовано вычисление относительной ошибки по каждой выходной переменной и усредненной ошибки по всем выходным переменным); для модифицированного алгоритма стохастической аппроксимации градиента – требуемая относительная ошибка; для генетических алгоритмов – максимальное количество ресурсов для алгоритма (количество индивидов, количество поколений) или удовлетворение требуемой относительной ошибки восстановления или прогнозирования зависимостей.

3.4.2 Выбор одного из четырех оптимизационных методов.

Алгоритм обратного распространения ошибки (задание скорости обучения), модифицированная процедура стохастической аппроксимации градиента целевой функции ошибки (настройки определены автором), самонастраивающийся ГА, гибридный ГА. Для решения сложных задач и при отсутствии априорной информации пользователю рекомендуется выбор генетических алгоритмов для обучения, настройки параметров, структур ИНС, пороговых значений. Первые два алгоритма функционируют в вещественном пространстве весовых коэффициентов, последние два - в бинаризованном пространстве весовых коэффициентов, структур ИНС, пороговых значений.

3.5 Запуск решения задач (и)

Кнопка «**Calculate**» запускает вручную процесс расчета с выбранными пользователем параметрами, кнопка «**Stop**» останавливает процесс расчета. Кнопки, в названии которых присутствует «**GA+Committees**», используют для обучения, настройки параметров и выбора структур ИНС, пороговых значений несколько алгоритмов и запускают вручную процесс расчета. По умолчанию в **ручном демонстрационном режиме** выполняется 1 запуск с выводом численных критериев эффективности и изменяющегося графика, отображающего процесс настройки ИНС, а также графика, отображающего изменение численных критериев.

В **автоматическом режиме** расчета, включаемом кнопкой «**Automatic mode**», выполняется перебор параметров на каждой задаче и множестве задач. Прогресс расчета отображается в трех строках. По умолчанию используется 500 запусков, что достаточно по результатам моих исследований и используемых генераторов псевдослучайных чисел для обеспечения необходимой статистической устойчивости и разброса значений численных критериев эффективности относительно математического ожидания. Выводятся результирующие графики после решения каждой задачи, основные результаты сохраняются в файл-отчет.

Осуществляется вывод графиков функций одной переменной, сечений графиков по задаваемым переменным для функций многих переменных, для которых осуществляется восстановление или прогнозирование. Результирующий выход настроенной ИНС на обучающей выборке обозначается темными точками, на экзаменующей выборке - светлыми точками. Также осуществляется вывод графиков изменения относительной ошибки по каждой переменной. В окне вывода результатов отображаются все основные критерии эффективности (усредняются по независимым прогонам алгоритмов): testing sample error for each output – ошибка ИНС на экзаменующей выборке для каждой выходной переменной; average testing sample error – ошибка ИНС на экзаменующей выборке, усредненная по всем выходным переменным; training sample

error for each output – ошибка ИНС на обучающей выборке для каждой выходной переменной; average training sample error – ошибка ИНС на обучающей выборке, усредненная по всем выходным переменным; number of criteria function calculations – число вычислений целевой оптимизируемой функции до выполнения условия останова обучения; number of algorithm iterations – число итераций алгоритма до выполнения условия останова обучения.

Список литературы

- [Звонков В.Б., 2011] Звонков В.Б. Разработка и применение системы IT-SAGA решения задач безусловной, условной однокритериальной и многокритериальной оптимизации при получении новых решений в региональных экономических задачах / В.Б. Звонков / Интеллектуальные системы и технологии: современное состояние и перспективы. Сборник научных трудов Международной летней школы-семинара по искусственному интеллекту для студентов, аспирантов и молодых ученых (Тверь-Протасово, 1-6 июля 2011 г.). – Тверь: Изд-во Тверского государственного технического университета, 2011 – с. 183-195.
- [Звонков В.Б., 2011] Звонков В.Б. Извлечение скрытых знаний и закономерностей из данных с использованием интеллектуальных алгоритмов обработки информации / В.Б. Звонков // Научный сборник. Материалы второй международной конференции «Автоматизация управления и интеллектуальные системы и среды», Нальчик: Издательство КБНЦ РАН, 2011, Т. II, секционные доклады – с. 117-121.
- [Звонков В.Б., 2012] Звонков В.Б. Многоагентные стохастические алгоритмы и комитеты интеллектуальных алгоритмов в задачах моделирования, оптимизации, прогнозирования и управления сложными системами (Multiagent stochastic algorithms and committees of intellectual algorithms for problems of modeling, optimization, forecasting and management of complex systems). // В.Б. Звонков / Сборник статей тринадцатой международной научно-практической конференции «Фундаментальные и прикладные исследования, разработка и применение высоких технологий в промышленности и экономике», 24-26 мая 2012 года, Санкт-Петербург, Россия, Издательство Политехнического университета, Том 1, с. 54-62.
- [Michalewicz, Z. et al., 1996] Michalewicz, Z. Evolutionary algorithms for constrained parameter optimization problems / Z. Michalewicz, M. Schoenauer // Evolutionary Computation. - 4. - 1996. - P. 1 – 32.
- [Michalewicz, Z. et al., 1996] Michalewicz, Z. Evolutionary algorithms for constrained engineering problems / Z. Michalewicz, D. Dasgupta, Riche Le, M. R. and M. Schoenauer // Computers & Industrial Engineering Journal - 30, 1996. – P. 851 – 870.
- [Zitzler E. et al, 1999] Zitzler E., Thiele L. Multiobjective evolutionary algorithms: A comparative case study and the strength Pareto approach // IEEE Transactions on Evolutionary Computation, Vol. 3, No. 4, pp. 257-271, 1999.