

ПРОГРАММНАЯ РЕАЛИЗАЦИЯ И ИССЛЕДОВАНИЕ ОСОБЕННОСТЕЙ
МЕТОДА ГРУППОВОГО УЧЕТА АРГУМЕНТОВ

Маркута О.В., Мысак В.Ф.

Стремительное развитие компьютерной техники существенно расширяет сферу использования методов математического моделирования для проектирования и управления сложными технологическими (техническими, экономическими, экологическими, финансовыми и др.) системами, прогнозирования результатов их деятельности, создания экспертно-диагностических подсистем и тренажерных комплексов по обучению персонала и т.п.

Эффективное использование методов моделирования требует простых и адекватных, как правило, формальных математических моделей. Одним из универсальных экспериментальных методов идентификации многомерных систем, применимым в режиме нормальной эксплуатации объектов исследования, есть метод группового учета аргументов (МГУА) [1,2]. Полученные по этому методу модели отображают неизвестные закономерности функционирования исследуемого объекта (вида “ n входов – 1 выход”), информация о котором неявно содержится в выборке экспериментальных данных. В МГУА для построения моделей применяются принципы автоматической генерации вариантов, промежуточных решений и последовательной селекции лучших моделей по определенным критериям. Алгоритм метода основан на делении выборки на части, при этом оценивание параметров и проверка качества моделей выполняются на разных подмножествах экспериментальной выборки. Эффективность метода подтверждена решением многочисленных реальных задач в таких областях как экология, гидрометеорология, экономика, техника [3].

Согласно МГУА решение трудоемкой задачи получения n -мерной (по входам) модели заменяется многостадийным процессом решения большого числа относительно простых задач аппроксимации экспериментальных данных функциями двух переменных. На каждой из стадий производится отбор наилучших полиномов, которые используются на последующих стадиях в качестве фиктивных аргументов новых полиномов.

Подобная процедура построения сложной функции от полинома продолжается до тех пор, пока на какой-либо из стадий не будет достигнута заданная точность аппроксимации.

Вместе с тем метод имеет ряд особенностей, не получивших строгого математического обоснования и требующих оперативного уточнения под конкретные условия использования получаемой модели. К таким особенностям относятся влияние на адекватность модели уровня шумов, количества отбираемых полиномов, размерности задачи, размеров подмножеств и т.п. Трудоемкая процедура таких уточнений требует использования современных программных средств моделирования.

В статье представлена программная реализация МГУА с генерированием исходных данных, ориентированная на широкий спектр возможных применений, и проведены исследования:

- определение возможных пределов работы приложения (максимального числа факторов);
- исследование влияния шума на точность полученной модели;
- исследование влияния количества отбираемых наилучших частных описаний на каждом шаге;
- исследование влияния размеров подмножеств экспериментальных данных на точность искомой модели;
- влияние вида частных описаний на качество полученной модели.

1. Программная реализация МГУА

Для исследования метода группового учета аргументов была выполнена его программная реализация. Разработанное приложение решает задачу построения модели объекта (вида “n входов – 1 выход”) по выборке экспериментальных данных.

На рис.1 показана общая блок-схема алгоритма работы приложения.

Алгоритмы МГУА воспроизводят схему массовой селекции. Входные аргументы и промежуточные переменные сопрягаются попарно, и сложность комбинаций на каждом ряду обработки информации возрастает (как при массовой селекции), пока не будет получена единственная модель оптимальной сложности [1].

Алгоритм работы программы:

Этап 1. Искусственная генерация исходных (экспериментальных) данных.

На этом этапе производится генерация значений векторов аргументов. При этом начальное количество аргументов (n) и размер выборки (число точек условного эксперимента) N изначально заданы. Количество аргументов n задается непосредственно, а размер выборки N определяется как сумма размеров трех ее подмножеств: обучающего (N1), проверочного (N2) и контрольного (N3):

$$N = N1 + N2 + N3 \quad (1)$$

Значения аргументов x_i генерируются в соответствии со следующим выражением:

$$x_{il} = \sin(\alpha_i \cdot N_l + \varphi_i), \quad i = \overline{1, n}; \quad l = \overline{1, N}, \quad (2)$$

где α_i и φ_i каждый раз генерируются псевдослучайным образом из множеств $[-A; A]$ и $[-F_i; F_i]$ соответственно, где значения A и F_i изначально задаются.

Полученные векторы аргументов содержат нормированные псевдослучайные значения.

Значения y_l генерируются следующим образом:

$$y_l = \hat{y}_l + y_l^m, \quad (3)$$

$$\hat{y}_l = b_0 + \sum_{i=1}^n b_i x_i + \sum_{\substack{i,j=1 \\ i \neq j}}^n b_{ij} x_i x_j + \sum_{i=1}^n b_{ii} x_i^2, \quad (4)$$

$$y_l^m = \frac{\hat{y}_l \cdot Sh_l}{100}, \quad l = \overline{1, N}, \quad (5)$$

где b_i и Sh_l каждый раз генерируются псевдослучайным образом из множеств $[-B; B]$ и $[-Sh; Sh]$, где значения B и Sh изначально задаются. Параметр Sh содержит значение уровня зашумления в % от величины значений y_l .

Этап 2. Реализация алгоритма МГУА.

Из множества аргументов $x=(x_1, x_2, \dots, x_n)$ выбираются пары аргументов (x_i, x_j) $i, j=1, n$; $i \neq j$. Для них составляются функции частных описаний квадратичного вида [1]:

$$y_{ij} = a_0 + a_i x_i + a_j x_j + a_{ij} x_i x_j + a_{ii} x_i^2 + a_{jj} x_j^2, \quad (6)$$

Используя метод наименьших квадратов (МНК) [4] для каждого описания по обучающему подмножеству $(\{y_u^o, x_{iu}^o, x_{ju}^o\} \in M_o, u = \overline{1, N_1})$ находится вектор неизвестных коэффициентов $a_{ij}^* = \{a_0, a_i, a_j, a_{ij}, a_{ii}, a_{jj}\}$.

По полученным моделям каждого частного описания вычисляются значения ординат (вектор \hat{y}_{ij}) для всех точек проверочного подмножества $\{x_{iu}^o, x_{ju}^o\} \in M_{II}, u = \overline{N_1, N_1 + N_2}$.

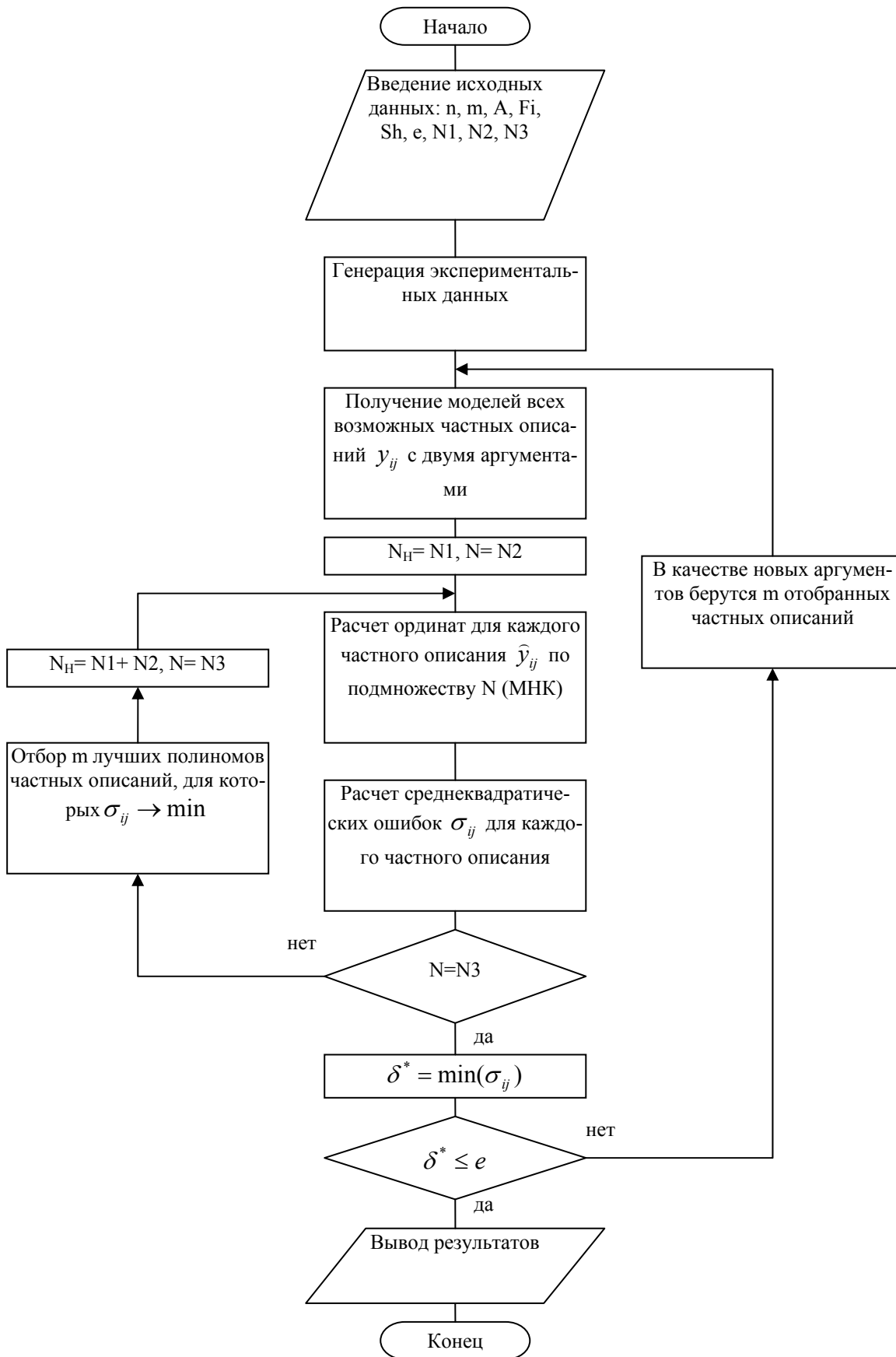


Рис. 1. Структурная блок-схема алгоритма разработанного приложения.

Далее производится расчет критерия селекции (среднеквадратической ошибки) для каждого частного описания по формуле:

$$\sigma_{ij} = \left\{ \frac{1}{N_2} \sum_{u=N_1+1}^{N_1+N_2} [y_u^{\circ} - \hat{y}_{ij}(u)]^2 \right\}^{0.5} \quad (7)$$

По критерию минимума на проверочном подмножестве отбираются m лучших моделей, т.е. реализуется процедура селекции.

Для отобранных моделей вычисляются значения ординат $\hat{y}_k, k = \overline{1, m}$ по контрольному подмножеству $\{x_{iu}^{\circ}, x_{ju}^{\circ}\} \in M_K, u = \overline{N_1 + N_2 + 1, N}$.

Далее для них вычисляются значения критерия селекции (так же по контрольному подмножеству):

$$\delta_k = \left\{ \frac{1}{N_3} \sum_{u=N_1+N_2+1}^N [y_u^{\circ} - \hat{y}_k(u)]^2 \right\}^{0.5} \quad (8)$$

Находится $\delta^* = \min(\delta_k)$.

Этап 3. Вывод результатов.

Производится вывод результатов в соответствующее поле окна приложения.

Выполняется проверка критерия останова:

$$\delta^* < \varepsilon \quad (9)$$

где $\varepsilon = \varepsilon$ (наперед заданная величина).

Если условие выполняется, то процесс поиска модели $\hat{y} = f(\bar{x}, a)$ считается завершенным. В противном случае производится возврат на начало *этапа 2* с предварительным пересчетом векторов аргументов (в качестве новых аргументов используются отобранные модели частных описаний).

2. Определение возможных пределов работы приложения (максимального числа факторов)

Число факторов (входов объекта) в принципе не ограничено. На практике работа приложения была проверена до $n = 25$. Но при таком количестве аргументов наблюдалось существенное увеличение времени работы программы (длительности всех необходимых расчетов). Если при $n=3, m=3$ время, затраченное на расчет одного шага алгоритма, составляло доли секунды, а загрузка ЦП составляла 11%, то при $n=25$ и $m=3$ время расчета одного шага увеличилось до 40 секунд, а загрузка ЦП – до 100%.

Вообще ограничения в отношении числа факторов могут накладываться только исходя из технических характеристик компьютера, на котором производится расчет модели программой. С увеличением числа факторов резко возрастает число их частных описаний и, следовательно, объем обрабатываемой информации, что в свою очередь вызывает увеличение требуемых ресурсов и приводит к существенному увеличению длительности расчетов.

3. Исследование влияния шума на точность полученной модели

При увеличении значения шума ухудшается точность искомой модели, как показано на рис.2. Это приводит к увеличению количества шагов, необходимых для нахождения модели, и к существенному ее усложнению.

Не во всех случаях удалось найти модели с заданной точностью аппроксимации. Некоторые из искусственно сгенерированных зависимостей $y = f(x_i)$ приняли вид, который трудно адекватно аппроксимировать квадратичным полиномом.

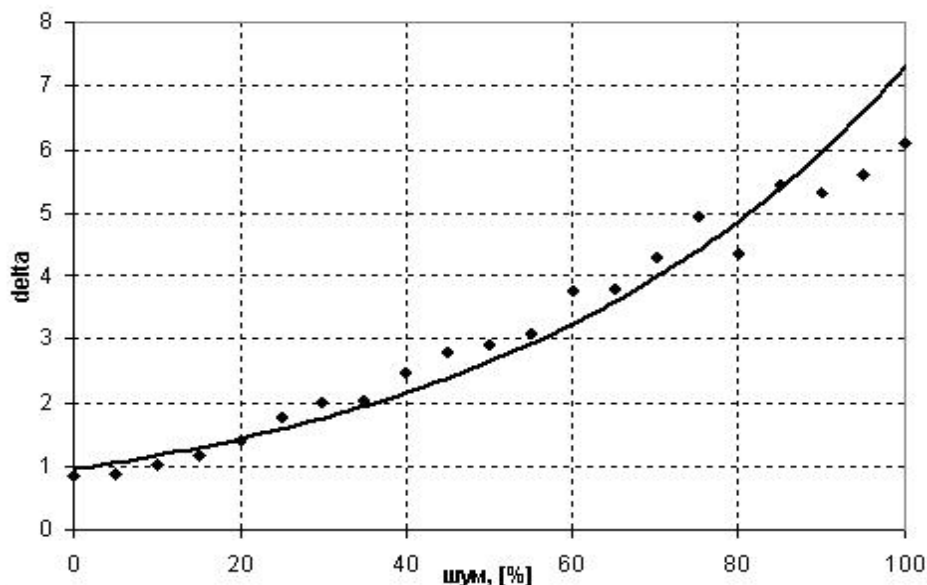


Рис. 2 Влияние уровня шума на точность полученной модели

4. Исследование влияния количества отбираемых наилучших частных описаний на каждом шаге

В плане выбора оптимального количества отбираемых на каждом шаге полиномов частных описаний однозначного ответа быть не может. Это обусловлено тем фактом, что с увеличением числа факторов существенно возрастает число их парных взаимодействий.

На практике прослеживается закономерность, что с увеличением числа отбираемых частных описаний на каждом шаге увеличивается точность полученной модели (рис.3). В свою очередь чрезмерное увеличение этого числа (m) приводит к большим объемам обрабатываемой информации.

Слишком же малое число отбираемых парных взаимодействий отрицательно сказывается на адекватности полученной модели.

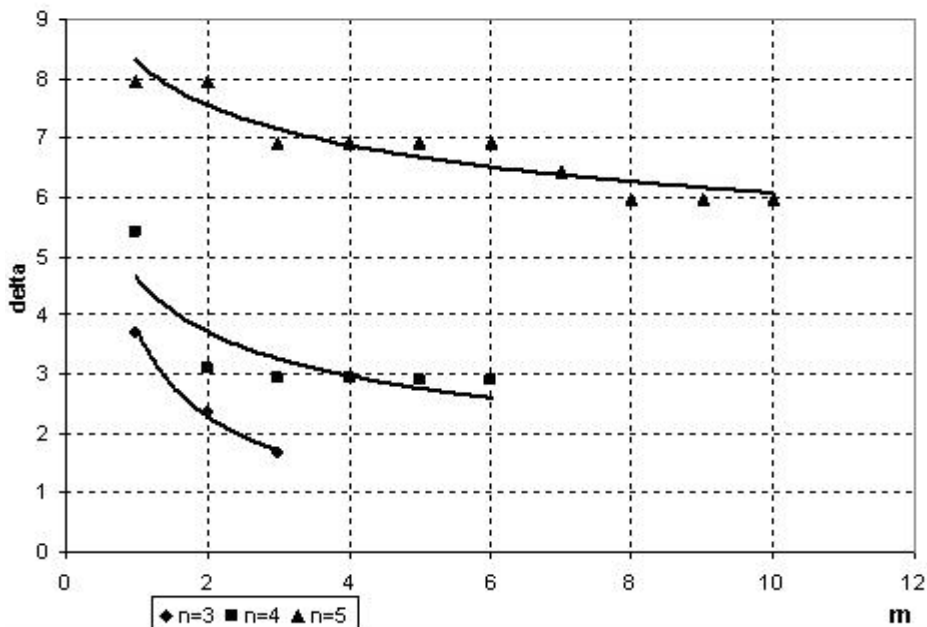


Рис. 3 Влияние числа отбираемых частных описаний на точность полученной модели для разных значений числа аргументов

Об оптимальном значении m можно говорить только в отношении конкретного n (числа исходных факторов). При решении конкретной задачи рекомендуется выполнить промежуточное исследование с использованием разработанного приложения для определения оптимального значения m .

5. Исследование влияния размеров подмножеств экспериментальных данных на точность искомой модели

В ходе исследования влияния размеров подмножеств O (обучающего), Π (проверочного) и K (контрольного) исходной выборки данных были получены следующие результаты:

1) При равномерном разбиении выборки ($O=\Pi=K$) точность модели прямо пропорциональна объему выборки (рис.4). Хотя чрезмерный объем данных может привести к существенному увеличению объема обрабатываемой информации.

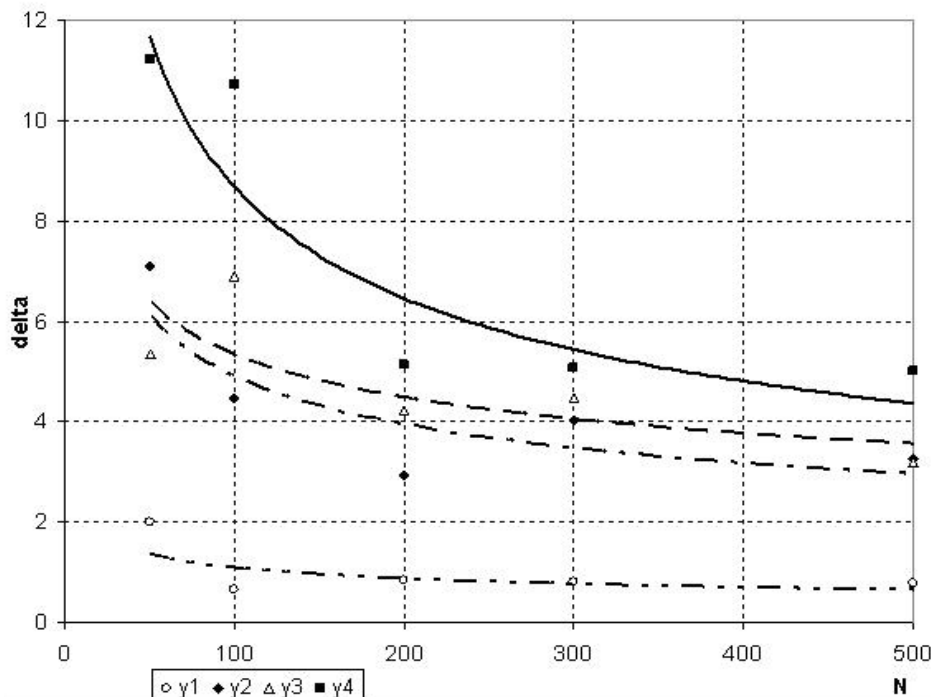


Рис. 4 Влияние размеров подмножеств экспериментальных данных на точность искомой модели при равномерном разбиении выборки ($O=\Pi=K$)

2) При неравномерном разбиении выборки было установлено, что увеличение обучающего подмножества по отношению к проверочному и контрольному ($O=\Pi+K$, $\Pi=K$) приводит к увеличению точности полученной модели по сравнению со случаем равномерного разбиения и разбиения в пользу Π либо K , как показано на рис.5.

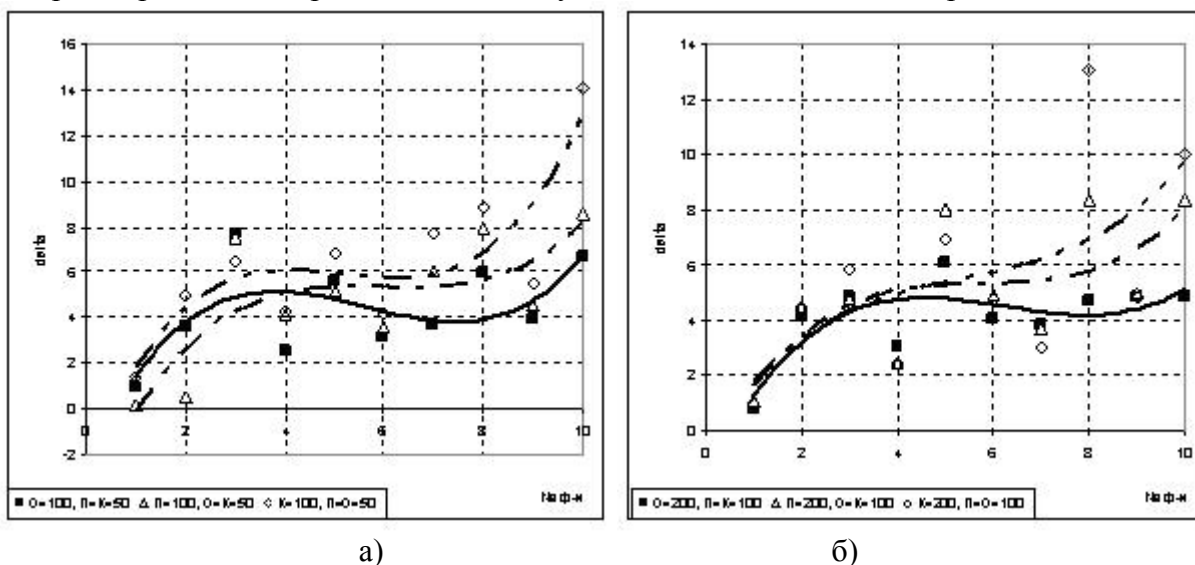


Рис. 5 Влияние размеров подмножеств экспериментальных данных на точность искомой модели при неравномерном разбиении выборки:
а) $N=200$; б) $N=400$

3) При включении в состав K дополнительно подмножеств O либо Π , либо $O+\Pi$ увеличивается точность полученной модели (изначально $O=\Pi=K$). Наилучшая точность была получена при $K=O+\Pi$ (рис.6), т.е. вся выборка делится на две части: обучающее и проверочное подмножества. А выбор оптимальной модели происходит при этом по всей выборке (всему множеству экспериментальных данных). Следующее по точности разбиение $K=O+\Pi+K$. Опять же при выборе модели учитывается вся выборка. Точность незначительно снижалась при разбиении $K=\Pi+K$.

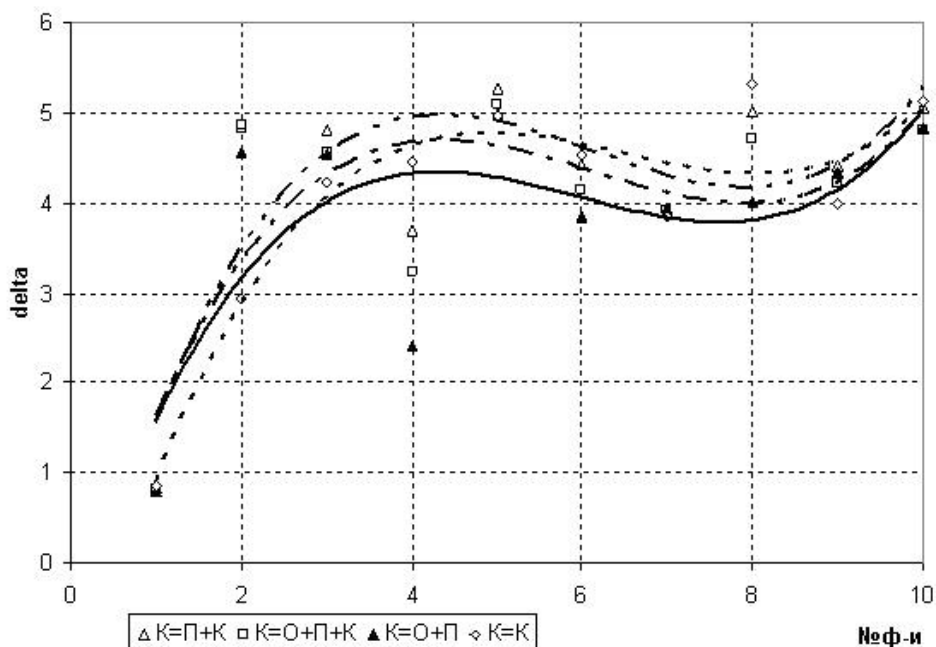


Рис. 6 Влияние включения в подмножество K дополнительных подмножеств экспериментальных данных на точность искомой модели (изначально $O=\Pi=K=200$)

Дополнительное включение в контрольное подмножество подмножеств O и Π повышает точность модели, поскольку в таком случае выбирается модель наиболее адекватная всей выборке. При стандартном разбиении (без включения подмножеств в состав K) есть риск, что на контрольном подмножестве функция, в целом не являющаяся наиболее точной моделью, будет иметь наименьшее среднеквадратическое отклонение от оригинала, что приведет к выбору модели не являющейся самой точной на всем исследуемом интервале экспериментальной выборки.

6. Влияние вида частных описаний на качество полученной модели

Изначально приложение реализовывало метод группового учета аргументов с функциями частных описаний квадратичного вида:

$$y_{ij} = a_0 + a_i x_i + a_j x_j + a_{ij} x_i x_j + a_{ii} x_i^2 + a_{jj} x_j^2. \quad (10)$$

В ходе исследований возник вопрос о влиянии вида полинома частных описаний на точность и сложность полученной модели.

Для сравнения были взяты функции частных описаний следующего вида:

$$y_{ij} = a_0 + a_i x_i + a_j x_j + a_{ij} x_i x_j. \quad (11)$$

В ходе исследований было установлено, что такое упрощение вида функций частных описаний практически не сказалось на качестве полученной модели (наблюдалось незначительное снижение точности, как показано на рис.7). При этом происходит существенное упрощение вида искомой модели (даже при возможном увеличении количества шагов расчета) при практически неизменной точности модели. При $n=3$, $m=3$, $e=0,75$ модель объекта в обоих случаях была найдена за два шага расчетов, но если для варианта с частными описаниями вида (10) модель содержала 64 слагаемых, а точность (δ) составила 0.46, то для случая с частными описаниями вида (11) модель уже содержала 16 слагаемых и $\delta=0.70$.

При некотором запасе точности использование упрощенного вида функций частных описаний может привести к существенному упрощению искомой модели.

Исследования проводились при следующих условиях: $n=3\dots 5$, $N_1=N_2=N_3=50\dots 500$, $m=2\dots 15$, $B=10$, $A=0.25$, $F_i=100$, $Sh=0.20$, $e=0.1\dots 5$.

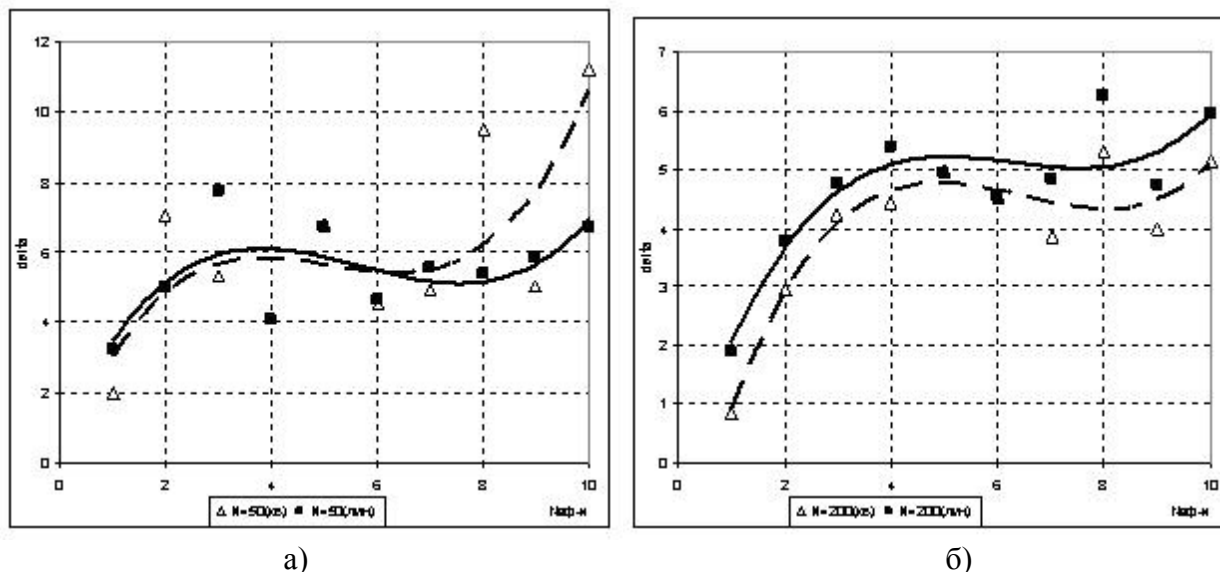


Рис. 7 Влияние вида частных описаний на точность искомой модели:
а) $N=50$; б) $N=200$

Выводы

На сегодняшний день одним из универсальных экспериментальных методов идентификации многомерных систем есть метод группового учета аргументов. Вместе с тем метод имеет ряд особенностей, уточнение которых являлось основной задачей проведенной работы.

В статье представлена программная реализация МГУА, ориентированная на широкий спектр возможных применений – от решения исследовательских задач до решения прикладных задач построения модели объекта (вида “ n входов – 1 выход”) по выборке экспериментальных данных. С помощью разработанного приложения возможно решение каждой конкретной задачи несколькими вариантами (при разных значениях параметров алгоритма) с целью получения модели объекта или системы оптимальной точности и сложности.

В ходе исследования практической (программной) реализации МГУА было установлено следующее:

1. С увеличением уровня шума и уменьшением числа отбираемых на каждом шаге частных описаний (m) ухудшается точность искомой модели, при этом возможно увеличение количества шагов, необходимых для нахождения модели, что может привести к существенному ее усложнению. При решении каждой конкретной задачи для определения оптимального значения m рекомендуется выполнить промежуточное исследование с использованием разработанного приложения.

2. При равномерном разбиении выборки ($O=P=K$) экспериментальных данных точность модели прямо пропорциональна объему выборки. При неравномерном разбиении выборки увеличение обучающего подмножества по отношению к проверочному и контрольному ($O=P+K$, $P=K$) приводит к увеличению точности полученной модели по сравнению со случаем равномерного разбиения и разбиения в пользу P либо K . При включении в состав K дополнительных подмножеств наилучшая точность была получена при $K=O+P$.

3. Вид частных описаний влияет на сложность полученной модели. Упрощение функций частных описаний от вида (10) к виду (11) приводит к упрощению модели, но при этом происходит незначительное ухудшение ее точности (при $n=3$, $m=3$, $e=0,75$ мо-

дель объекта в обоих случаях была найдена за два шага расчетов, но если для варианта с частными описаниями вида (10) модель содержала 64 слагаемых, а точность (δ) составила 0.46, то для случая с частными описаниями вида (11) модель уже содержала 16 слагаемых и $\delta=0.70$).

При решении каждой отдельной задачи, для нахождения модели с оптимальными параметрами сложности и точности, рекомендуется поварьировать с использованием разработанного программного продукта такие параметры как число отбираемых частных описаний на каждом шаге и вид полинома частных описаний.

In the article the program application of the method of the group calculation of the arguments is represented. The following special features of method were investigated with the aid of the developed application: influence on the adequacy of the model of the level of noise, quantity of selected polynomials, dimensionality of task, sizes of subsets. The results of studies were represented.

1. Ивахненко А. Г. Долгосрочное прогнозирование и управление сложными системами. – Киев: Техніка, 1975.
2. Ивахненко А.Г. Индуктивные методы самоорганизации моделей сложных систем. – Киев: Наук. мысль, 1982.
3. Степашко В.С. Теоретические аспекты МГУА как метода индуктивного моделирования. – Труды I Международной конференции по индуктивному моделированию, Львов, 20-25 мая 2002 г.
4. Демидович Б.П., Марон И.А., Шувалова Э.З. Численные методы анализа. Приближение функций, дифференциальные и интегральные уравнения. – М.: Наука, 1967.

УДК 004.02:681.3

МОДЕЛЬ АВТОМАТИЗОВАНОЇ СИСТЕМИ УПРАВЛІННЯ ПОТОКАМИ ТРАНСПОРТНИХ ЗАСОБІВ

Славич В.П.

Постановка проблеми. Постійне збільшення кількості транспортних засобів призводить до перевантаженості доріг, росту транспортних затримок та появи тривалих пробок. Невпинне зростання інтенсивностей транспортного руху викликає необхідність в розробці методів, спрямованих на збільшення пропускних можливостей дорожньої мережі та зменшення відповідних наслідків існуючих транспортних проблем.

Сучасний розв'язок транспортних труднощів полягає в організації автоматизованого управління дорожнім рухом та розробці відповідної бази. Важливою складовою такої автоматизованої системи управління транспортними потоками є інформаційно-математичні методи і моделі, що забезпечують раціональне управління об'єктом транспортної мережі. Завдяки отриманню інформації про параметри потоку в реальному часі, а також враховуючи статистичні параметри процесу, подібна модель використовуються для формування керуючих дій, що передаються на засоби дорожньої сигналізації [1, 2] та безпосередньо впливають на транспортний потік.

Аналіз існуючих методів управління транспортною системою дає висновок, що всі подібні моделі мають два основні недоліки. Перший недолік – це спрощеність моделі, другий – вузьке застосування. Отже, розв'язок, запропонований ними, має невелике практичне значення [3-7].

В зв'язку з цим виникає задача розробки методів управління потоками автомобілів, які б забезпечували прийнятний результат.