

ТЕХНОЛОГИЯ РЕШЕНИЯ ЗАДАЧИ МОДЕЛИРОВАНИЯ И ПРОГНОЗИРОВАНИЯ НА ОСНОВЕ ИНДУКТИВНОГО ПОДХОДА

Е.А. Савченко

Международный научно-учебный центр информационных технологий и систем НАН Украины и МОН Украины

Разработана и описана технология решения задачи моделирования и прогнозирования на основе индуктивного подхода. В ее основе лежит комбинаторный алгоритм МГУА и метод доопределения. Технология позволяет автоматически, по выборке экспериментальных данных, находить модель объекта или процесса с использованием двух внешних критериев выбора: точности и несмещенности. Применение технологии в реальных задачах моделирования и прогнозирования подтвердило ее эффективность.

Ключевые слова: индуктивный подход, комбинаторный алгоритм метода группового учета аргументов, моделирование, прогнозирование, технология.

Розроблено та описано технологію розв'язання задачі моделювання та прогнозування на основі індуктивного підходу. В її основі лежить комбінаторний алгоритм МГУА та метод довизначення. Технологія дозволяє автоматично, за вибіркою експериментальних даних, знаходити модель об'єкту або процесу з використанням двох зовнішніх критеріїв вибору: точності і незміщеності. Застосування технології в реальних задачах моделювання та прогнозування підтвердило її ефективність.

Ключові слова: індуктивний підхід, комбінаторний алгоритм методу групового урахування аргументів, моделювання, прогнозування, технологія.

ВВЕДЕНИЕ

Задача построения моделей, которые описывают или прогнозируют поведение наблюдаемого объекта или процесса, является неотъемлемой составной частью решения задач управления и принятия решений. Для построения математической модели в общем случае следует определить ее структуру и оценить параметры, то есть решить задачу структурно-параметрической идентификации или моделирования по выборке статистических или экспериментальных данных.

Часто реальные прикладные задачи характеризуются нестабильностью измерений, трудно учесть все факторы, влияющие на процесс. Для моделирования и прогнозирования таких данных целесообразно использовать индуктивный подход как метод построения моделей по данным наблюдений. Индуктивные методы позволяют автоматически находить зависимости, скрытые в выборке экспериментальных данных.

Один из основных алгоритмов индуктивного моделирования — комбинаторный алгоритм метода группового учета аргументов (МГУА). Однако применение этого алгоритма в реальных задачах показало, что не всегда удается однозначно определить модель по одному заданному критерию. Для такого случая разработан метод доопределения модели на

основе комбинаторного алгоритма МГУА.

Цель работы — разработать технологию решения задачи моделирования и прогнозирования по выборке экспериментальных данных на основе комбинаторного алгоритма МГУА и метода доопределения, которая позволяла бы строить оптимальную модель в реальных прикладных задачах.

ОСОБЕННОСТИ ИНДУКТИВНОГО ПОДХОДА

Индуктивное моделирование базируется на переборе множества моделей-кандидатов по внешним критериям селекции и решает задачу структурно-параметрической идентификации. Одним из наиболее эффективных методов моделирования по экспериментальным данным в условиях неполноты информации является метод группового учета аргументов, автором которого является академик А.Г. Ивахненко [1–2].

В отличие от регрессионного анализа, где структура модели задается, в МГУА структура оптимальной модели и ее параметры находятся посредством самоорганизации моделей, то есть испытания многих моделей-кандидатов по внешним критериям селекции. В [3] проводилось сравнение эффективности применения регрессионного анализа и МГУА для прогнозирования экономических процессов, где сравнивались прогнозирующие свойства моделей, построенных на основе метода наименьших квадратов (МНК) и МГУА, на примерах анализа объема производства легкой промышленности и инфляции. Было показано, что прогнозирующие свойства модели МГУА значительно выше, причем модель оптимальной сложности значительно проще, чем полная модель, построенная по МНК. Некоторые факторы, которые не вошли в модель, построенную по МГУА, были не просто лишними и малоинформативными, а также «вредными» в условиях имеющейся короткой выборки данных.

МГУА построен на принципах самоорганизации, а именно, на основе следующих двух принципов.

1. При постепенном усложнении структуры значения внешних критериев сначала уменьшаются, а затем увеличиваются, то есть в их зависимости от сложности моделей существует минимум; только внешние критерии, рассчитанные по данным, которые не использовались для построения моделей, имеют такой минимум. Поэтому в МГУА применяется деление не менее чем на две независимые выборки данных [1]. Одна из них используется для оценки параметров, а вторая — для оценки качества модели. Применять только один критерий принципиально нельзя, поскольку может возникнуть ситуация: «чем сложнее модель, тем она точнее». Все критерии, которые используются в МГУА, основаны на разбиении выборки, т.е. на «внешней» информации, поэтому названы внешними критериями.

2. Особенностью МГУА является обеспечение «свободы выбора». При постепенном усложнении структуры моделей на каждом ряду из лучших по внешним критериям моделей отбирается не одна, а несколько лучших моделей [2]. Наиболее известным среди методов индуктивного моделирования является комбинаторный алгоритм МГУА [4]. Это метод,

который не требует доказательства сходимости, поскольку выполняет полный перебор всех возможных моделей-кандидатов. МГУА разработан для решения задач моделирования сложных систем, прогнозирования, идентификации и аппроксимации многофакторных систем, диагностики, распознавания образов и кластеризации данных.

ПОСТРОЕНИЕ ТЕХНОЛОГИИ МОДЕЛИРОВАНИЯ И ПРОГНОЗИРОВАНИЯ НА ОСНОВЕ ИНДУКТИВНОГО ПОДХОДА

Применение комбинаторного алгоритма МГУА в прикладных задачах показало, что возникает необходимость в повышении эффективности выбора оптимальной модели, поскольку часто вместо оптимальной модели избирается множество моделей-кандидатов с близкими значениями внешнего критерия регулярности. Поэтому на основе комбинаторного алгоритма МГУА был разработан метод поиска оптимальной модели — метод доопределения с применением критерия несмещенности [5].

Это двухкритериальный метод доопределения с использованием критерия несмещенности ошибок, который позволяет устранить неоднозначность выбора модели в реальных задачах моделирования за счет применения целесообразной последовательности критериев селекции. Исследование различных последовательностей внешних критериев МГУА показало, что следующая последовательность является целесообразной [6]: сначала применяется критерий точности, а затем критерий несмещенности. Такое последовательное применение двух внешних критериев повышает помехоустойчивость модели.

Индуктивное моделирование при самоорганизации моделей выдвигает определенные требования к выборке исходных данных. Во-первых, выборка должна содержать те аргументы, которые влияют на исходную переменную. Способы выбора информативных аргументов приведены в [7]. Во-вторых, выборка должна быть полной, то есть не содержать пропущенных значений [8]. Если данные в выборке имеют разную размерность, желательно их нормировать.

Разработана компьютерная технология моделирования и прогнозирования, которая включает следующие этапы [9]:

- 1) предварительная обработка выборки данных;
- 2) построение модели идентификации или прогноза.

На рис. 1 показана общая схема компьютерной технологии моделирования и прогнозирования.

Выборка данных (1) подается на вход блока предварительной обработки данных (2), после чего поступает на вход комбинаторного алгоритма МГУА (3), на выходе которого получаем оптимальную модель (4) или множество моделей-кандидатов, которые имеют минимальное значение внешнего критерия. В случае необходимости при построении модели осуществляется доопределение выбора модели с целью отбора оптимальной модели.

Блок 1 — *Исходная выборка данных* представляет собой матрицу экспериментальных данных, содержащую n наблюдений за ходом процесса

(строк выборки), m входных переменных (столбцов выборки) и вектор n наблюдений выходной переменной.

Блок 2 — Предварительная обработка выборки данных.

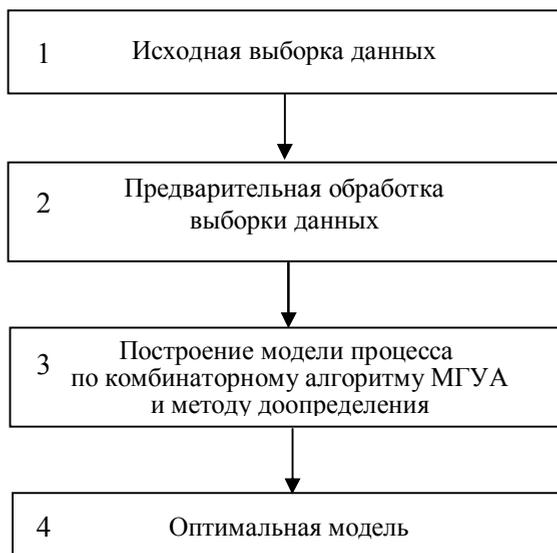


Рис. 1. Построение прогнозирующей модели с использованием комбинаторного алгоритма МГУА и метода доопределения

Предварительная обработка выборки данных включает следующие задачи [8]:

- восстановление пропущенных данных наблюдений и расширение состава аргументов за счет дополнительных;
- нормирование исходных данных;
- расширение выборки данных за счет введения дополнительных аргументов;
- выбор наиболее информативных аргументов;
- оптимизация размера выборки за счет поиска аналогов в предыстории.

Рассмотрим этапы этого блока подробнее.

1. Сначала выборка проверяется на наличие пропущенных значений. Для восстановления пропущенных данных в выборке данных применяется комбинаторный алгоритм МГУА с различными шаблонами считывания данных [9]. По заданному шаблону формируется расширенная выборка данных, по которой определяется лучшая модель и рассчитывается значение пропущенного элемента. Для каждого пропуска в данных рассчитывается новая модель.

2. Следующим шагом является нормирование данных, которое целесообразно применять только тогда, когда данные имеют разную размерность или очень разные масштабы чисел.

3. Далее формируются дополнительные аргументы. Расширение пространства моделирования за счет введения таких аргументов может повысить точность полученной модели. К данным исходной выборки

добавляются попарные произведения начальных переменных, запаздывания и тому подобное [7].

4. Оценка информативности входных и сформированных переменных осуществляется с помощью расчета значений модуля коэффициента корреляции каждой переменной с выходной переменной. Переменные, которые имеют малое значение модуля, могут быть исключены из множества входных аргументов.

5. Для оптимизации размера выборки рассчитывается значение евклидовых расстояний вектор-строк таблицы данных, а затем отбирается из всего множества наблюдений необходимое количество ближайших к исходному наблюдению.

Блок 3 — Построение модели процесса по комбинаторному алгоритму МГУА и методу доопределения.

Рассмотрим подробнее структуру комбинаторного алгоритма МГУА с доопределением по критерию несмещенности ошибок. Сначала все наблюдения ранжируются по значениям дисперсии:

$$D_j^2 = \sum_{i=1}^m (x_{ij} - \bar{x}_i)^2,$$

где $j = 1, 2, \dots, m$, m — число аргументов,

$$\bar{x}_i = \frac{1}{n} \sum_{j=1}^n x_{ij},$$

где n — общее число точек.

Для определения оптимальной модели выборки должны как можно больше отличаться [1]. Поэтому исходная выборка делится на две части: первые из всех наблюдений, которые имеют наибольшее значение дисперсии, составляют обучающую подвыборку, оставшаяся часть всех наблюдений — проверочную подвыборку (с наименьшим значением дисперсии). Для всех генерируемых в процессе работы алгоритма моделей различной сложности рассчитываются значения критерия регулярности. По минимуму этого критерия выбирается множество лучших моделей, для каждой модели рассчитывается значение критерия несмещенности ошибок и по его минимуму выбирается оптимальная модель (*Блок 4*).

На рис. 2 показана процедура доопределения по критерию несмещенности ошибок.

На основе разработанной технологии разработан комплекс программ, который позволяет решать задачи предварительной обработки выборки данных для моделирования и прогнозирования. Комплекс программ, ядром которого является комбинаторный алгоритм МГУА, реализован на языке программирования C++.

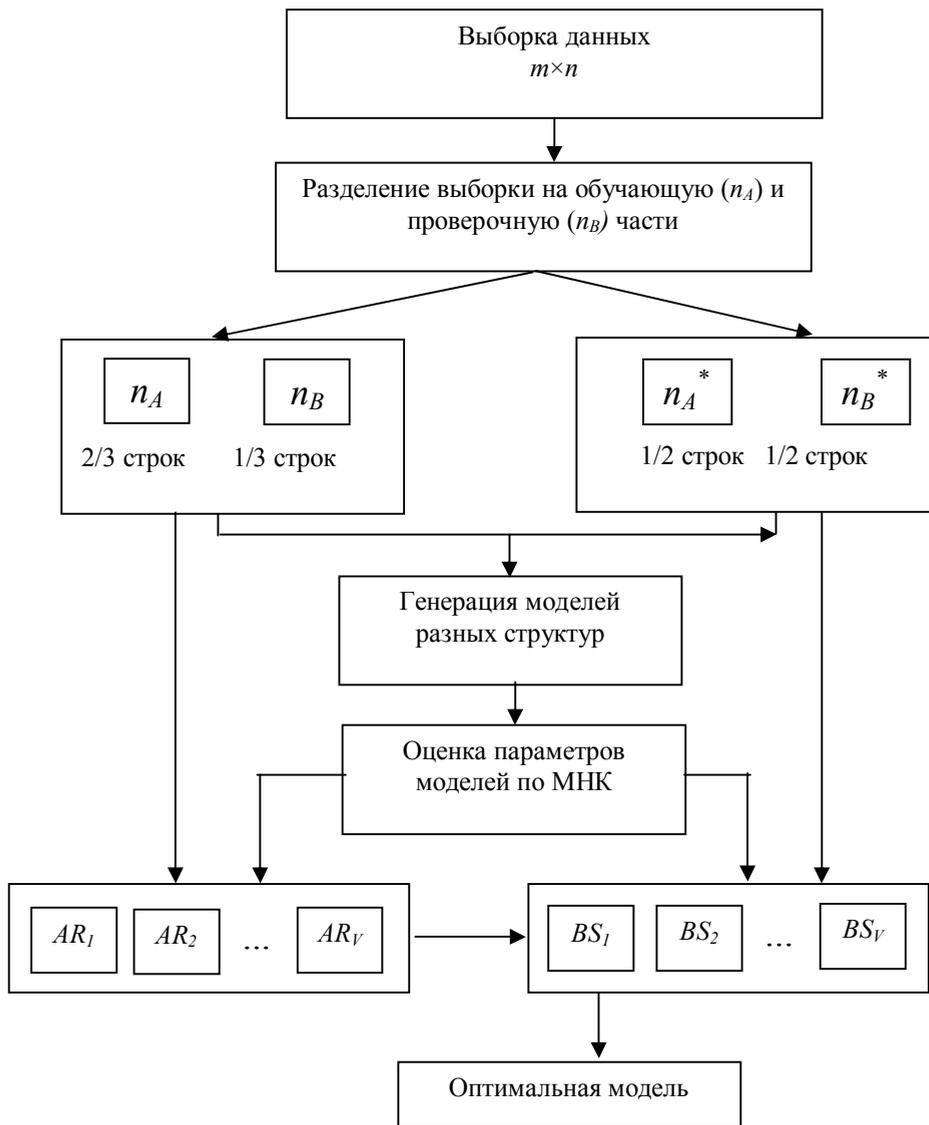


Рис. 2. Процедура доопределения по критерию несмещенности ошибок:

n_A — длина обучающей части выборки данных; n_B — длина проверочной части выборки данных; AR_1, AR_2, \dots, AR_V — значения критерия регулярности для лучших моделей, которые попали в интервал неопределенности; BS_1, BS_2, \dots, BS_V — значения критерия несмещенности для лучших по критерию регулярности моделей.

ПРИМЕНЕНИЕ ТЕХНОЛОГИИ МОДЕЛИРОВАНИЯ И ПРОГНОЗИРОВАНИЯ ДЛЯ РЕШЕНИЯ ПРАКТИЧЕСКИХ ЗАДАЧ

Проведено исследование эффективности описанной технологии на основе метода доопределения для выбора заданных моделей на тестовых примерах. Примеры показали, что полученная в результате работы данной

технологии модель является наиболее точной не только на обучающей и проверочной частях выборки, но и на третьей, независимой, экзаменационной части [10].

Разработанная компьютерная технология апробирована при решении прикладных задач моделирования и прогнозирования: в задачах лечения диабета в условиях надомного мониторинга и задачах моделирования взаимодействия ионов струи газов с поверхностью материалов.

В задаче лечения диабета по данным надомного мониторинга проведено восстановление отсутствующих значений и получен краткосрочный прогноз уровня глюкозы в крови. Пример восстановления единичных пропусков исходных данных с различными шаблонами показал, что наиболее точными для восстановления единичных пропусков является крестообразный и квадратный шаблоны [9]. Выполнено сравнение результатов восстановления с МГУА и по методу экспоненциального сглаживания и показана эффективность первого метода.

При моделировании взаимодействия ионов струи газов с поверхностью материалов задача состояла в подборе материала покрытия поверхности космического аппарата, который содержит наименьший коэффициент распыления [11]. Совместно с учеными Московского авиационного института на основе экспериментальных данных, содержащих значения коэффициента распыления ряда материалов ионами ксенона при фиксированных параметрах газа, найдена зависимость коэффициента распыления от характеристик этих материалов. С помощью описанной технологии получена модель, которая имеет полиномиальную структуру, и хотя она выбрана не из физических соображений, но с помощью самоорганизации модели по внешнему критерию, оказалось, что она потенциально способна описать систему точнее, чем традиционные модели с физической параметризацией. Ошибка полученной модели около 20–40%, в отличие от моделей с физической параметризацией, которые могут давать ошибку намного больше.

Выводы

Разработана и описана компьютерная технология, которая обеспечивает эффективное решение задач моделирования и прогнозирования по экспериментальным данным, что подтверждено тестовыми примерами. Эта технология обеспечивает повышение помехоустойчивости моделей за счет последовательного применения внешних критериев МГУА: критерия регулярности и несмещенности.

Применение разработанной технологии в решении реальных задач моделирования и прогнозирования показало ее эффективность.

1. Ивахненко А.Г. Помехоустойчивость моделирования / А.Г. Ивахненко, В.С. Степашко. — Киев : Наукова думка, 1985. — 215 с.
2. Madala H.R. Inductive Learning Algorithms for Complex Systems Modeling / H.R. Madala, A.G. Ivakhnenko — Boca Raton: CRC Press Inc., 1994. — 384 p.

3. Ивахненко А.Г. Метод группового учета аргументов — конкурент метода стохастической аппроксимации / А.Г. Ивахненко // Автоматика. — 1968. — № 3. — С. 58–72.
4. Степашко В.С. Комбинаторный алгоритм МГУА с оптимальной схемой перебора моделей / В.С. Степашко // Автоматика. — 1981. — № 3. — С.31–36.
5. Ивахненко А.Г. Исследование эффективности метода доопределения выбора модели в задачах моделированием с применением МГУА / А.Г.Ивахненко, Е.А. Савченко // Проблемы управления и информатики. — 2008. — № 2. — С. 65–76.
6. Савченко Е.А., Степашко В.С. Аналитическое и численное исследование селективных свойств критерия несмещенности ошибок в задачах индуктивного моделирования / Е.А. Савченко, В.С. Степашко // Проблемы управления и информатики. — № 2. — 2012. — С. 39–50.
7. Ивахненко А.Г. Концепция последовательных алгоритмических приближений (спусков) к точному решению интерполяционных задач искусственного интеллекта / А.Г. Ивахненко, Г.А. Ивахненко, Е.А. Савченко // Кибернетика и вычислительная техника. — № 124. — 1999. — С. 40–60.
8. Савченко Е.А. Предварительная обработка выборки данных в задаче индуктивного моделирования / Е.А. Савченко // Управляющие системы и машины. — № 2. — 2015. — С. 82–87.
9. Ивахненко А.Г. Применение алгоритмов МГУА для восстановления пропущенных данных и прогноза уровня глюкозы в крови при надомном мониторинге диабета / А.Г. Ивахненко, Е.А. Савченко, Г.А. Ивахненко, Т. Гергей // Проблемы управления и информатики. — 2002. — № 3. — С. 123–133.
10. Савченко Е. А. Анализ селективных свойств критериев МГУА при их последовательном применении / Е.А. Савченко, В.С. Степашко // Моделивання та керування станом еколого-економічних систем регіону. Збірник праць. — Київ: МННЦ ІТС, 2008. — С. 199–210.
11. Ивахненко А.Г. Обнаружение закономерностей взаимодействия ионов с поверхностью по комбинаторному алгоритму МГУА / А.Г. Ивахненко, Е.А. Савченко, Г.А. Ивахненко [та ін.] // Проблемы управления и информатики. — 2003. — № 2. — С. 80–89.

UDC 681.513

TECHNOLOGY FOR SOLVING THE PROBLEM OF MODELING AND FORECASTING BASED ON INDUCTIVE APPROACH

E.A. Savchenko

*International Research and Training Center for Information Technologies and
Systems of National Academy of Sciences of Ukraine and Ministry of Education and
Science of Ukraine*

Introduction. The advantage of inductive algorithms is in their ability to automatically find dependencies hidden in a sample of experimental data. Combinatorial algorithms of GMDH (group method of data handling) are the main inductive modeling algorithms. These algorithms applied to real problems showed that it's not always possible to unambiguously determine a model by one criterion. Method of a model after-determination based on the Combinatorial GMDH algorithm is developed for such case. A technology based on the combinatorial GMDH algorithm and the after-determination method was developed for the modeling and forecasting.

The purpose of this article is to develop the technology for modeling and forecasting on the experimental data sample based on the combinatorial algorithm

GMDH method and the after-determination method. This technology will help to find the optimal model in real applications.

Results. A technology for solving the problem of modeling and forecasting on the basis of the inductive approach was developed and described. This approach is based on the combinatorial algorithm GMDH method and completions. This technology, based on a sample of experimental data, automatically finds the object model or process using two external selection criteria: accuracy and bias. The developed computer technology was tested in solving applied problems of modeling and prediction: in problems of diabetes in a home-based monitoring and problems of modeling the interaction of ions with the surface of the jet gas materials.

Conclusion. A computer technology that provides an effective solution for the problems of modeling and prediction of the experimental data was developed and described. Numerical examples demonstrate its efficiency. This technology provides increased noise immunity models due to the consistent application of external criteria GMDH: the criterion of regularity and bias. This technology was used in real applications for modeling and forecasting and its effectiveness has been confirmed.

Keywords: inductive approach, combinatorial algorithm of group method of data handling, modeling, forecasting, technology.

1. Ivakhnenko A.G. *Noise-immunity of modeling*. Kiev: Naukova Dumka, 1985, 215 p. (in Russian).
2. Madala H.R. *Inductive Learning Algorithms for Complex Systems Modeling*. Boca Raton: CRC Press Inc., 1994, 384 p.
3. Ivakhnenko A.G. Group method of data handling as competitor to the method of stochastic approximation. *Soviet Automatic Control*, 1968, no. 3, pp. 64–78.
4. Stepashko V.S. Combinatorial GMDH algorithm with optimal scheme of models sorting. *Automatics*, 1981, no. 3, pp. 31–36 (In Russian).
5. Ivakhnenko A.G. Investigation of Efficiency of Additional Determination Method of the Model Selection in the Modeling Problems by Application of the GMDH Algorithm. *Journal of Automation and Information sciences*. Begell House: Inc. Publishers, 2008, vol. 40, no. 3, pp. 47–58.
6. Savchenko E.A. Analytical and Numerical Study of the Selective Properties of the Errors Unbiasedness Criterion in the Problems of Inductive Modeling. *Journal of Automation and Information sciences*. Begell House: Inc. Publishers, 2012, vol. 44, no. 4, pp. 1–12.
7. Ivakhnenko A.G. Conception of the successive algorithmic approaching (lowering) to the exact decision of interpolation tasks of artificial intelligence. *Cybernetics and computing engineering*, 1999, vol. 124, pp. 40–60 (In Russian).
8. Savchenko E.A. Preprocessing of data sample in inductive modeling problem. *Control Systems and Computers*, 2015, N2, pp. 82–87.
9. Ivakhnenko A.G. Application of Algorithms of the Method of Batch Assessment of Arguments for Recovering Missed Data and Prediction of the Glucose Level in Blood on at Home Diabetes Monitoring. *Journal of Automation and Information sciences*. Begell House: Inc. Publishers, 2002, vol. 34, no. 6, pp.123–133.
10. Savchenko E.A. Analysis of the selective properties of the GMDH criteria when applying them consistently. *Modeling and control as ecological and economic systems of the region*. Kyiv: IRTC ITS, 2008, no. 4, pp. 199–210 (In Russian).

Получено 25.03.2015