

Исследование эффективности нечёткого группового метода обработки данных в задачах прогнозирования

Федяев О.И., Гладунов С.А.

Кафедра ПМИ, ДонГТУ

fedyayev@r5.dgtu.donetsk.ua

Abstract

Fedyayev O., Gladunov S. Analysis of efficiency of fuzzy group method of data processing in tasks of prediction. In the work there are discussed tasks of constructing and using of fuzzy models, used to raise precision of planning and conducting complex objects, information about which is incomplete, distorted and conflicting. There are results of modeling described on the example of prediction of behavior of processes with different nature.

Введение

В работе рассматриваются вопросы построения и применения нечетких моделей, используемых для повышения достоверности планирования и управления сложными объектами, сведения о которых неполны, зашумлены и противоречивы. К таким объектам (задачам) относятся оценка природных и техногенных рисков, биржевые и финансовые предсказания, моделирование кризисных ситуаций, распознавание образов и т.п. Анализ и последующий прогноз событий, порождаемых подобными процессами (явлениями), может выполняться тремя способами:

- традиционными детерминированными математическими методами (в общем случае это построение и решение дифференциальных уравнений);
- статистическими методами;
- с применением методов искусственного интеллекта (экспертные системы, нейронные сети, средства нечеткой математики, генетические алгоритмы).

Математическое моделирование, основанное на строгом алгоритмическом описании объекта (процесса), является методом, адекватным крайне узкому кругу задач. Для решения подавляющего большинства проблем управления, возникающих в природных системах и сложных объектах жизнедеятельности человеческого общества, требуются методы, способные учитывать нечеткость, ассоциативность, неопределенность и другие аналогичные характеристики реальной жизни. Одним из таких методов является групповой метод обработки данных (ГМОД) для моделирования сложных многомерных нелинейных систем. Он основан на принципах эвристической самоорганизации по наблюдаемым данным на входе и выходе системы [1]. В этом методе параметры модели отождествляются с нечеткими числами, а моделирование выполняется с позиции нечетких явлений и систем.

1. Математические основы ГМОД

Данная статья является продолжением работы авторов по исследованию эффективности нечеткого ГМОД в задаче прогнозирования поведения сложных систем, математическая постановка которой приведена в статье [2].

Метод основан на сглаживании таблично заданной функции полиномом Колмогорова - Габора [1]:

$$Y = R_0 + \sum_0^n R_n + \sum \sum R_{n_1 n_2} X_{n_1} X_{n_2} + \sum \sum \sum R_{n_1 n_2 n_3} X_{n_1} X_{n_2} X_{n_3} + \dots ,$$

где $Y=X(T)$; $X_i =X(T-i)$ - некоторая предыстория протекания процесса в момент времени T ; R_i - коэффициенты полинома.

Нахождение коэффициентов возможно двумя методами:

- решением системы нормальных уравнений Гаусса;
- при помощи алгоритмов самоорганизации.

В первом случае использование нормальных уравнений - достаточно трудоемкая задача, поскольку уже при четырех аргументах она требует решения системы порядка 70×70 .

Другой подход, предложенный А.Г. Ивахненко [1], позволяет существенно снизить объем вычислений с сохранением уровня адекватности (достоверности) модели. В случае с четырьмя аргументами задача сводится к нахождению коэффициентов трёх полиномов:

$$Y_1 = R_{01} + R_{11}X_1 + R_{21}X_2 + R_{31}X_1^2 + R_{41}X_2^2 + R_{51}(X_1X_2 + X_1X_3 + X_2X_4),$$

$$Y_2 = R_{02} + R_{12}X_3 + R_{22}X_4 + R_{32}X_3^2 + R_{42}X_4^2 + R_{52}(X_3X_4 + X_1X_4 + X_2X_3),$$

$$Y = R_1 + R_1Y_1 + R_2Y_2 + R_3Y_1^2 + R_4Y_2^2 + R_5Y_1Y_2,$$

т.е. к решению трёх систем порядка 6×6 , коэффициенты которых можно находить, например, методом наименьших квадратов. При этом вычисляются Y_1 и Y_2 как функции от времени, на их основе получают коэффициенты R_0, \dots, R_5 , и, путем обратной замены, получается окончательная запись полинома Колмогорова - Габора.

Остановимся на ГМОД подробнее. В этом методе при построении модели процесса применяются принципы биологической селекции. Принцип селекции заключается в "скрещивании" переменных (генов) и выборе наиболее удачных моделей (особей) на своем уровне итераций (поколении).

На первом этапе для каждой пары входных векторов строится модель процесса по формуле

$$Y_k = R_1 + R_{k1}X_i + R_{k2}X_j + R_{k3}X_i^2 + R_{k4}X_j^2 + R_{k5}X_iX_j,$$

получая таким образом $N(N-1)/2$ пар, где N - количество входных переменных. Из этих моделей получаем новый набор переменных. Из него выбираем некоторые наиболее близкие к исходному процессу вектора (селекция) и строим новые модели на их основе. Процесс продолжается до тех пор, пока не перестанет уменьшаться некоторый критерий качества, например, степень близости наилучшей модели данного уровня к исходному процессу. По окончании итерационного процесса производится обратная замена промежуточных переменных, в результате чего получается окончательная модель.

2. Особенности нечеткой реализации ГМОД

Все процессы можно условно разделить на детерминированные и вероятностные. К детерминированным относятся процессы, протекание которых можно предсказать с абсолютной точностью в любой момент времени. Для вероятностных процессов нельзя дать точный прогноз, можно лишь рассчитать интервал, в который с той или иной вероятностью попадет заданная величина.

Поэтому было бы целесообразно использовать для прогнозирования подход, работающий с интервальными величинами.

Под интервалом будем понимать величину, характеризуемую минимальным (a-c) и максимальным (a+c) значениями, а также некоторой функцией принадлежности. В настоящей работе рассматриваются интервальные величины с линейной функцией принадлежности (рис. 1), т.е. случайные величины с равномерным законом распределения.

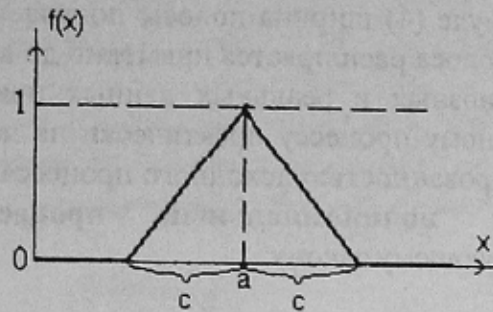


Рис. 1. Функция принадлежности нечеткого числа

Учитывая целесообразность нечеткого подхода в задачах прогнозирования и вероятностных оценок, интерес вызывает возможность распространить его на различные методы прогнозирования. В настоящей работе исследуется возможность и эффективность построения нечетких моделей сложных процессов методом групповой обработки данных.

3. Анализ результатов прогнозирования с помощью ГМОД

Оценка результатов прогнозирования проводилась на динамических процессах различной природы: гармонической синусоидальной, гармонической экспоненциальной и случайной.

Гармонический синусоидальный процесс. Динамика изменения процесса задавалась формулой:

$$f(t)=2+\sin(10*t*\pi/50)+\epsilon, \tag{4}$$

где t - модельное время, изменяющееся в диапазоне от 1 до 100; ϵ - случайная равномерно распределенная величина, принимающая значения в интервале $[-0,1; 0,1]$.

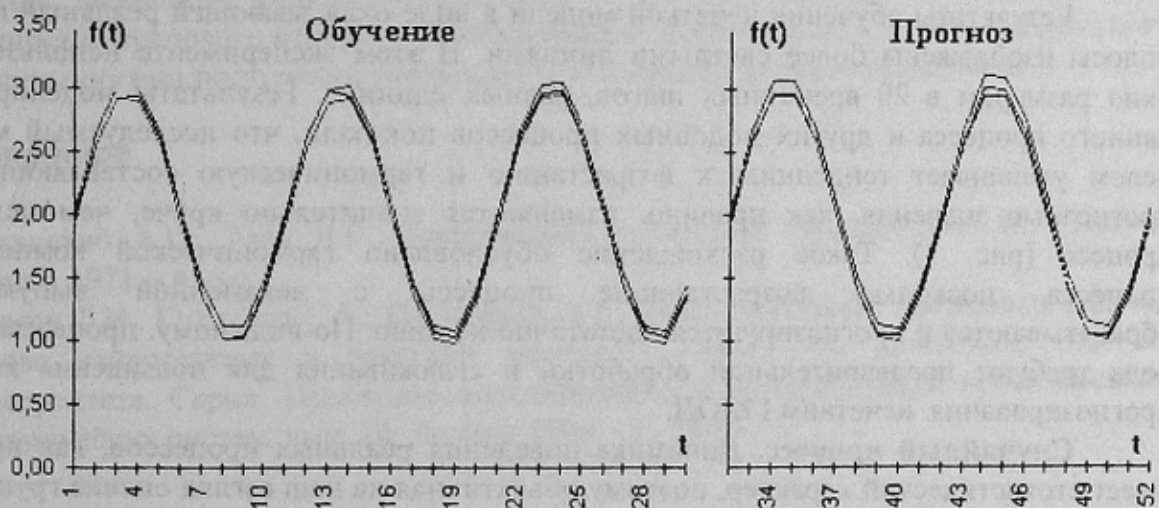


Рис. 2. Результаты прогнозирования гармонического синусоидального процесса с помощью нечёткой модели

На рис. 2 представлены результаты обучения нечёткой модели и прогнозирования динамики процесса на её основе. Для обучения использовались

данные, формируемые методом окон [2]. В этом эксперименте использовалось окно размером 20 единиц модельного времени. Оптимальная модель данного процесса получалась на шестом уровне селекции. Она позволяет построить интервальную полосу, охватывающую реальный процесс на всем обучающем множестве. Ширина полосы зависит от природы исходных данных о процессе. При отсутствии случайной составляющей ϵ в формуле (4) ширина полосы получается нулевая, а при её наличии, как видно из рисунка, полоса расширяется примерно до величины помехи.

Сравнение прогнозных и реальных данных показало хорошую адекватность нечеткой модели реальному процессу практически на любом интервале времени. Это обусловлено детерминированностью исходного процесса.

Гармонический полиномиальный процесс. Данные о процессе формировались по следующему закону:

$$f(t) = e^{\frac{t}{20}} \cdot |\sin(t)|,$$

причем t изменялось в интервале $[0; 100]$. Графически этот процесс показан темной линией на рис. 3.

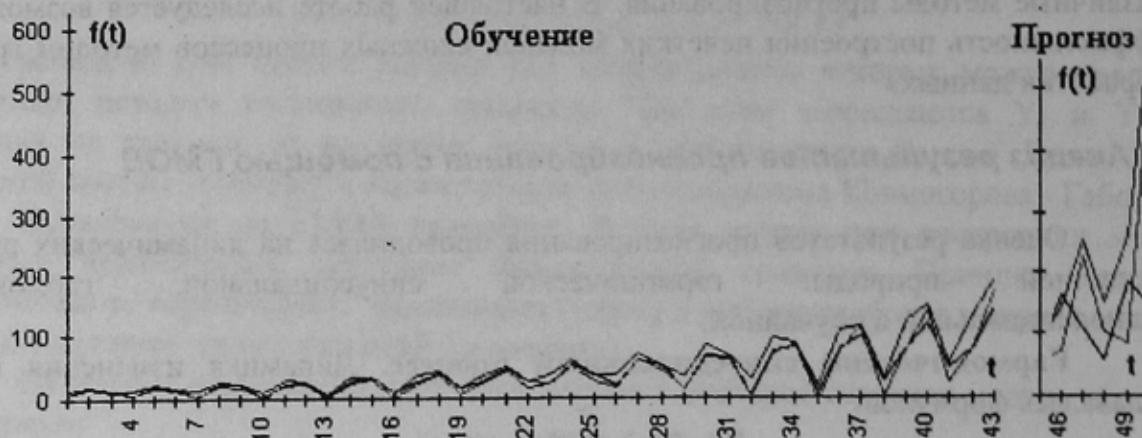


Рис. 3. Результаты прогнозирования гармонического экспоненциального процесса с помощью нечеткой модели

Результаты обучения нечеткой модели в виде охватывающей реальный процесс полосы изображены более светлыми линиями. В этом эксперименте использовалось окно размером в 20 временных шагов, равных единице. Результаты моделирования данного процесса и других подобных процессов показали, что исследуемый метод в целом улавливает тенденцию к возрастанию и гармоническую составляющую, но прогнозные значения, как правило, изменяются значительно круче, чем исходный процесс (рис. 3). Такое расхождение обусловлено гармонической компонентой процесса, поскольку возрастающие процессы с неизменной выпуклостью обрабатываются и прогнозируются достаточно хорошо. По-видимому, процессы такого рода требуют предварительной обработки и сглаживания для повышения качества прогнозирования нечетким ГМОД.

Случайный процесс. Динамика поведения реальных процессов, как правило, имеет стохастический характер, поэтому объективная на наш взгляд оценка группового метода была получена при обработке случайного процесса, в качестве которого брались данные, описывающие среднемесячный сток реки Северский Донец за 17 лет (204 значения).

Данные о процессе изменялись циклически с периодом в 12 месяцев. Легко заметить, что весной сток реки был максимальным, а зимой достигал минимума. В

целом изменение данных соответствует гармоническому синусоидальному закону, исследованному выше, но при этом содержит значительный элемент случайности в значениях амплитуды и форме пиков стока воды. Эта особенность процесса в виде графика показана на рис. 4 более темной линией. Результаты прогнозирования, представленные на рис. 4 светлыми линиями, показывают, что нечеткая модель процесса, полученная с помощью ГМОД, оказалась достаточно адекватной, так как позволила довольно точно воспроизвести структуру процесса на 16 месяцев вперед. Как видно из рис. 4, основной ошибкой в прогнозе оказалось смещение пика с апреля на май, но величина выброса была предсказана достаточно точно, поскольку наиболее вероятное значение предсказанного процесса находится в середине интервальной полосы, нижняя часть которой на графике практически совпала с осью времени.

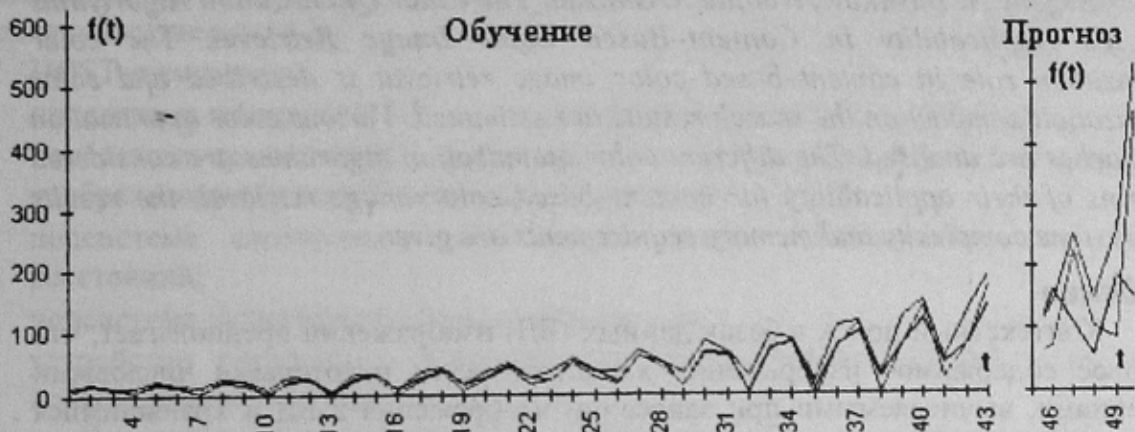


Рис. 4. Результаты прогнозирования помесячного стока реки Северский Донец

Заключение

Проведенные исследования позволяют сделать вывод, что нечеткий групповой метод обработки данных наиболее эффективен при прогнозировании гармонических процессов с горизонтальной линией регрессии. Благодаря нечеткой логике метода удается построить интервальную модель, позволяющую оценить диапазон отклонений от тренда, вызванных воздействием случайной составляющей процесса. Отмеченные особенности группового метода обработки данных обосновывают целесообразность его использования при построении реальных систем анализа и прогнозирования.

Литература

1. Ивахненко А.Г., Лапа В.Г. Предсказание случайных процессов. - К.: "Наукова думка", 1971 - 416 с.
2. Федяев О.И., Гладунов С.А., Прокофьев А.В. Прогнозирование временных рядов на основе нейросетевых и нечетких моделей. - Науч. тр. Донецкого гос. тех. университета. Серия: Проблемы моделирования и автоматизации проектирования динамических систем. вып. 10, Донецк 1999 - С. 38-43.