

АВТОМАТИЗАЦИЯ ПРОГНОЗИРОВАНИЯ ВРЕМЕННЫХ РЯДОВ ГЕНЕТИЧЕСКИМ АЛГОРИТМОМ

Паротькин Н.Ю.

Научный руководитель – доцент, к.т.н. Жуков В.Г.

*Сибирский государственный аэрокосмический университет имени академика
М.Ф. Решетнёва*

Анализ временных рядов играет важную роль в решении многих актуальных задач. Чаще всего для этой цели применяется аппарат математической статистики, что обязывает использующего его человека разбираться в применяемых моделях и понимать заложенный в них смысл, поскольку использование неверной модели дает результат далекий от реальности. Это обусловлено тем, что выполнение численных операций в различных моделях не приводит к алгебраическим ошибкам и, следовательно, применение ошибочной модели незаметно для неквалифицированного человека. Использование статистических пакетов так же не позволяет избежать описанной проблемы, поскольку в них автоматизировано вычисление по выбранной модели и выполняется формализованная проверка входных данных. Кроме того данные программные продукты обладают сложным пользовательским интерфейсом с возможностью настройки большого количества параметров мало знакомых обычному пользователю.

Чаще всего обозначенной группе пользователей необходим краткосрочный прогноз на один – два шага вперед на основе временного ряда содержащего 50 – 100 значений, поскольку накопление большей статистики по интересующему пользователя процессу накапливается редко в связи с отсутствием необходимости в ней. Следовательно, для данной категории пользователей необходимо использовать модель краткосрочного прогнозирования процессов с преобладающей детерминированной составляющей, представленных временными рядами с короткой длиной актуальной части, в условиях отсутствия априорной информации о вероятностных характеристиках процесса с автоматической настройкой параметров модели для достижения оптимальных результатов. Перечисленным требованиям удовлетворяет модель временного ряда на основе нечетких множеств, оптимальные параметры которой будут настраиваться при помощи генетического алгоритма (ГА).

Целью данной работы является реализация модели нечеткого временного ряда, её настройка при помощи ГА и сравнение результатов прогнозирования с данными полученными на основе моделирования при помощи нейронной сети. Данный выбор обусловлен тем, что от пользователя при использовании данных методов не требуется знание математической статистики и умения правильно выбрать статистическую модель для временного ряда.

Приведем пошаговое описание алгоритма построения нечеткого временного ряда.

1. **Определение универсума.** Пусть $f_i(t)=T_i$, ($i=1,2,\dots$) – реальные значения временного ряда для некоторого фактора. Определим универсум U для приращений значений фактора как $U = [D_{min} - D_1, D_{max} + D_2]$, где D_{min} ($D_{min} = \min_t(f(t) - f(t - 1))$) и D_{max} ($D_{max} = \max_t(f(t) - f(t - 1))$) – минимальное и максимальное приращения значений фактора на основе известных данных временного ряда

соответственно, а D_1 и D_2 – два действительных числа, использование которых позволяет разбить универсум U на n интервалов равной длины: u_1, u_2, \dots, u_n .

2. **Фаззификация.** Выполним фаззификацию данных, определив интервалы $u_r (r = \overline{1, n})$, которым принадлежит приращение фактора. Если значение приращения фактора принадлежит интервалу u_1 , то соответствующее ему нечеткое значение имеет вид: $X_1 = 1/u_1 + 0,5/u_2$. Если значение приращения фактора принадлежит интервалу u_r , то соответствующее ему нечеткое значение имеет вид: $X_r = 0,5/u_{r-1} + 1/u_r + 0,5/u_{r+1}$, $r = \overline{1, n-1}$. Если значение приращения фактора принадлежит интервалу u_n , то соответствующее ему нечеткое значение имеет вид: $X_n = 0,5/u_{n-1} + 1/u_n$.

3. **Формирование логических отношений $X_k \rightarrow X_j$.** Пусть X_k и X_j – нечеткие значения фактора для i -го и $(i+1)$ - периодов соответственно. Тогда для i -го периода можно записать нечеткую логическую зависимость вида: $X_k \rightarrow X_j$. Нечеткое импликативное отношение будем интерпретировать как импликация Мамдани, т.е. нечеткое импликативное отношение $D = B \rightarrow C$ для произвольных векторов B и C вычисляется по формуле $d_{ij} = b_i^T \times c_j = \min(b_i, c_j)$.

4. **Вычисление отношений R_k .** Определим нечеткие логические зависимости для всех известных значений временного ряда, а затем объединим нечеткие логические зависимости с одинаковой левой частью в одну группу. Так, если были сформированы нечеткие зависимости $X_k \rightarrow X_j, X_k \rightarrow X_l, X_k \rightarrow X_s$, то они объединяются в группу $X_k \rightarrow X_j, X_l, X_s$. Результирующее отношение R_k представляет собой объединение логических отношений, попавших в k -тую группу, т.е.

$$R_k = X_k^T \times X_j \cup X_k^T \times \dots \times X_l \cup X_k^T \times X_s \quad (1)$$

5. **Прогнозирование и дефаззификация получаемых результатов.** Результирующее НМ для прогнозируемого значения временного ряда для $(i+1)$ -го периода вычисляется по формуле: $X_{i+1} = X_i \circ R_i$, где X_i – фаззифицированное приращение предшествующего периода, R_i – нечеткое отношение для i -го периода, вычисленное по формуле (3). Дефаззификация полученного НМ осуществляется по методу центра тяжести для одноточечных множеств

$$y_{i+1} = \sum_{r=1}^n c_r w_r / \sum_{r=1}^n c_r, \quad (2)$$

где n – количество интервалов $u_r (r = \overline{1, n})$, w_r – средняя точка r -го интервала, c_r – значение степени принадлежности для r -го интервала результирующего НМ, описывающего группу нечетких зависимостей.

6. **Вычисление прогнозируемой величины.** Искомое значение прогнозируемой величины находится как сумма реального значения временного ряда фактора T_i для i -го периода и дефаззифицированного (четкого) значения приращения фактора y_{i+1} , рассчитываемого по формуле 2 или 3.

Учет повторяющихся НМ в правых частях нечетких логических групп позволит повысить точность прогноза. Если при формировании групп нечетких зависимостей были выявлены повторяющиеся НМ в правых частях групп, то значение приращения для $(i+1)$ -го периода:

$$y_{i+1} = \sum_{r=1}^n v_r w_r / \sum_{r=1}^n v_r, \quad (3)$$

где n – количество интервалов $u_r (r = \overline{1, n})$, w_r – дефаззифицированное значение (по формуле (4) или (5)), v_r – количество повторений НМ X_r в правой части нечеткой логической зависимости, описывающей прогноз для $(i+1)$ -го периода.

Для оценки качества прогноза может быть использована средняя относительная ошибка прогноза (AFER):

$$AFER = \frac{\sum_{i=1}^m |(F_i - T_i)/T_i|}{m} \cdot 100\%, \quad (4)$$

где F_i и T_i – предсказанное и реальное значения для i -го периода, m – количество значений временного ряда. И относительная ошибка прогнозирования:

$$error = \frac{|F_{i+1} - T_{i+1}|}{T_{i+1}} \cdot 100\%, \quad (5)$$

где F_{i+1} и T_{i+1} – предсказанное и реальное значения для $(i+1)$ -го периода.

Для обеспечения возможности работы обычного пользователя с приведенной выше моделью необходимо автоматизировать поиск оптимальных значений для действительных чисел D_1 , D_2 , используемых при корректировке универсума U , количества интервалов разбиения n универсума U и параметра $\alpha \in [0,1]$ вместо степени принадлежности 0,5 в лингвистических терминах u_r . Применение ГА позволит автоматизировать нахождение оптимальной комбинации значений выше перечисленных параметров. Поэтому хромосома ГА будет иметь следующий вид: (D_1, D_2, n, α) .

Количество бит и диапазон значений для D_1 и D_2 определяется пользователем, но предустановленным значением для них будет являться принадлежность интервалу $[0, d]$, где $d = |D_{max} - D_{min}|$. Для $n - [2, n_{max}]$, где $n \in \mathbb{N}, n_{max} \leq m - 1 \leq 33$ (m – количество значений временного ряда), соответственно 5 бит. $\alpha \in [0,1]$, 7 бит.

В качестве целевой функции может быть использована функция (4) и её следует минимизировать. Во время работы ГА может получиться комбинация параметров, при которой при вычислении по формулам (2), (3) получится значение вида “0/0” из-за наличия группы логических нечетких зависимостей с неопределенными правыми частями. Для исключения данных комбинаций из дальнейшего рассмотрения необходимо назначить им максимально худшее значение функции пригодности – 100.

Выше описанный алгоритм построения нечеткого временного ряда и ГА, позволяющий находить его оптимальные параметры, были реализованы в виде программного средства.

Для проверки работоспособности реализованного алгоритма проведем прогнозирование и моделирование с его помощью 4 временных рядов. Их графики приведены на рис. 1.

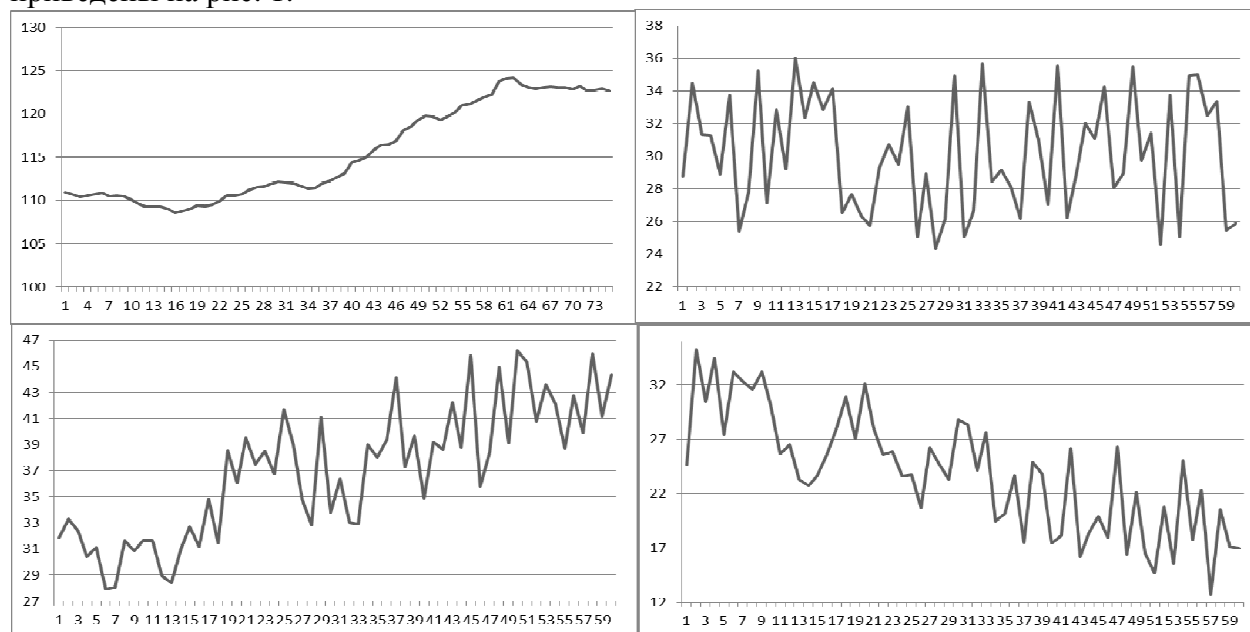


Рисунок 1 Графики временных рядов

Максимально возможные значения параметров для НВР приводятся в таблице 1. По результатам тестирования были найдены оптимальные параметры НВР для каждого из четырех рядов. Они так же приведены в таблице 1. Усредненные значения по 20 запускам программы для каждого ряда для средней относительной ошибки прогноза (4) и относительной ошибки прогнозирования (5) приводятся в таблице 2.

Таблица 1 Параметры НВР для тестовых временных рядов

№ ряда		D_1	D_2	n	α
1	Значение	[0; 1,533]	[0; 1,533]	20	[0;1]
	Количество бит	14	14	5	7
	Наилучшее значение	0,39831	0,43827	7	0,95313
2	Значение	[0; 13,199]	[0; 13,199]	30	[0;1]
	Количество бит	13	13	5	7
	Наилучшее значение	4,45176	5,46682	2	0,95313
3	Значение	[0; 12,235]	[0; 12,235]	30	[0;1]
	Количество бит	13	13	5	7
	Наилучшее значение	2,32095	9,40476	3	0,4375
4	Значение	[0; 13,698]	[0; 13,698]	30	[0;1]
	Количество бит	13	13	5	7
	Наилучшее значение	8,96339	8,05042	2	0,42188

Для сравнения эффективности интеллектуальных алгоритмов, при прогнозировании временных рядов, рассмотренные ранее тестовые данные были обработаны с помощью нейронной сети. Перед подачей данных на вход нейронной сети к ним был применен метод обработки данных – скользящее окно. Глубина окна была выбрана равной 3. Для всех временных рядов структура и настройки сети были одинаковыми. Входной слой был представлен 3 нейронами, так же один скрытый слой из двух нейронов и один нейрон в выходном слое. В качестве активационной функции была использована сигмоида. Точность аппроксимации и прогнозирования так же оценивалась с помощью формул (4), (5) и приведена в таблице 2.

Таблица 2 Ошибки прогнозирования НВР и нейросети

№ ряда	НВР		Нейросеть	
	AFER	error	AFER	error
1	0,24898%	0,23415%	0,23427%	0,19793%
2	14,53275%	1,60655%	7,58778%	12,98713%
3	8,53093%	2,94043%	5,95831%	2,85008%
4	18,18118%	2,37419%	11,99008%	13,92147%

Результаты проведенных исследований показывают, что использование нечетких временных рядов совместно с генетическим алгоритмом позволяет создать программное средство обладающее минимум настраиваемых параметров для достижения оптимальных показателей надежности прогнозирования при стандартных настройках. Рассматриваемая модель подходит для краткосрочного прогнозирования как стационарных, так и нестационарных коротких временных рядов, что подтверждается низким значением ошибки на тестовых данных.

В сравнении с другим интеллектуальным методом анализа данных: нейросетью – данная комбинация алгоритмов незначительно уступает ей в точности аппроксимации и значительно лучшие в точности прогнозирования. При этом следует учитывать, что использованная для тестирования нейронная сеть обладала простейшей структурой. Данный момент определялся тем, что неподготовленный пользователь самостоятельно выберет наиболее простой вариант и поэтому он эквивалентен настройке параметров в комбинации НВР и ГА.

Следовательно, рассмотренная в статье комбинация алгоритмов и созданная на их основе программа предназначена для пользователей, мало знакомых с математической статистикой, при этом желающих получить достоверный краткосрочный прогноз на основе незначительного количества статистических данных.