

Скворцова Т.С. Анализ и разработка нечеткой модели с использованием генетического алгоритма для прогнозирования временных рядов. // Проблемы информатики в образовании, управлении, экономике и технике: Сб. статей VIII Всерос. научно-техн. конф. – Пенза: ПДЗ, 2008. – С. 252-254.

АНАЛИЗ И РАЗРАБОТКА НЕЧЕТКОЙ МОДЕЛИ С ИСПОЛЬЗОВАНИЕМ ГЕНЕТИЧЕСКОГО АЛГОРИТМА ДЛЯ ПРОГНОЗИРОВАНИЯ ВРЕМЕННЫХ РЯДОВ

Т.С. Скворцова

Рязанский государственный радиотехнический университет,
г. Рязань

Прогнозирование будущих значений временного ряда на основании предыдущих данных играет важную роль в решении многих актуальных задач, например, прогноз температуры. Большинство реальных событий характеризуются некоторой неопределенностью. Поэтому каждому наблюдению нечеткого временного ряда можно поставить в соответствие нечеткую переменную с некоторой функцией принадлежности.

Нечеткие временные ряды могут быть представлены с помощью нечетких множеств. Пусть $F_1(t)$ и $F_2(t)$ – нечеткие временные ряды. Зависимость

$$(F_1(t-k), F_2(t-k)), \dots, (F_1(t-2), F_2(t-2)), (F_1(t-1), F_2(t-1)) \rightarrow F_1(t), \quad (1)$$

где $F_1(t)$ – главный фактор прогнозирования, $F_2(t)$ – вспомогательный фактор прогнозирования, называется двухфакторной моделью прогноза k -го порядка на основе нечетких временных рядов.

Область значений главного фактора прогнозирования определяется как $U = [D_{\min} - D_1, D_{\max} + D_2]$, где D_{\min} и D_{\max} – минимальное и максимальное значения исторических данных главного фактора прогнозирования, а D_1 и D_2 – два действительных числа, использование которых позволяет разбить универсум U на n интервалов u_1, u_2, \dots, u_n равной длины. Аналогично определяется вспомогательный фактор на универсуме V .

При фаззификации данных определяются интервалы u_i ($i = 1, 2, \dots, n$), которым принадлежат значения главного фактора. Если значение фактора принадлежит интервалу u_i , то нечеткое значение

$$X_i = 0,5/A_{i-1} + 1/A_i + 0,5/A_{i+1},$$

где $A_i = 0/u_1 + 0/u_2 + \dots + 0,5/u_{i-1} + 1/u_i + 0,5/u_{i+1} + \dots + 0/u_{n-1} + 0/u_n$.

По полученным данным для каждого интервала временного ряда строится нечеткое логическое отношение. Прогнозирование выполняется по методу, предложенному в [1]. При этом средняя относительная ошибка прогноза (AFER) может быть вычислена по формуле

$$AFER = \frac{\sum_{i=1}^m |(F_i - T_i)/T_i|}{m} \cdot 100\%, \quad (2)$$

где F_i и T_i – предсказанное и реальное значения для i -го периода; m – количество значений временного ряда.

Отдельная задача заключается в выборе основных параметров предложенной модели (значения D_1, D_2, E_1, E_2 , число интервалов разбиения n и m , порядок модели k) таким образом, чтобы ошибка прогнозирования была минимальной.

Применение классического генетического алгоритма позволяет сократить время выбора оптимальных параметров и улучшить результат прогнозирования. При этом хромосома будет иметь вид $S_1 = (D_1, D_2, E_1, E_2, n, m, k)$.

В качестве функции соответствия можно выбрать функцию (2), видоизмененную следующим образом. Если для некоторой хромосомы определены все правые части групп нечетких логических зависимостей, то функция соответствия определяется по формуле (2). Если не определена хотя бы одна правая часть, то значение функции соответствия $J(s)$ находится как сумма средней относительной ошибки прогноза по формуле (2) и числа 100.

$$J(s) = \begin{cases} AFER, & \text{если определены все правые части,} \\ AFER + 100, & \text{иначе,} \end{cases} \quad (3)$$

где $AFER$ определяется по формуле (3).

Хромосома, которая обеспечивает минимум целевой функции по формуле (3), будет признанной лучшей. Выбор родителя будет состоять в определении лучшей хромосомы, которая минимизирует функцию (3), из двух случайно выбранных. Затем две найденные таким образом хромосомы-родителя будут использоваться для скрещивания. При выполнении скрещивания выбирается коэффициент скрещивания R_c и генерируется случайное число $N_c = random([0,1])$. Если $R_c > N_c$, то случайным образом выбирается точка скрещивания z , и выполняется скрещивание. Мутация выполняется аналогичным образом.

Метеорологическое прогнозирование является неотъемлемой частью современной жизни. Обычно такой прогноз представляет собой глобальную задачу, включающую сбор метеоданных на большой территории за длительный промежуток времени, изучение их эволюции и только потом прогноз. Предложенный выше метод позволяет на основании обработки исторических данных, наблюдаемых в определенном месте, прогнозировать будущие значения.

На основе двухфакторных моделей с использованием генетического алгоритма был выполнен анализ температуры в июне 2007 г. в городе Камбарка в Удмуртской Республике. В качестве прогнозируемого фактора была выбрана температура, в качестве вспомогательного – влажность. Были определены следующие параметры модели прогнозирования: числа корректировки диапазона $D_1 = -0,745101$, $D_2 = 0,184253$, $E_1 = -0,596672$, $E_2 = 1,109230$, количество интервалов разбиения для каждого фактора $n = 11$, $m = 4$, порядок модели $k = 3$. Интервал изменения прогнозируемого фактора получается: $[5,6449; 24,2743]$. Очевидно, что без использования ГА подбор этих значений занял бы очень много времени. Ошибка прогнозирования, вычисленная по формуле (3), составила 2,4335%.

Библиографический список

1. Lee, L.W., Wang, L.H., Chen, S.M. Handling forecasting problems based on two-factors high-order fuzzy time series // *IEEE Transactions on fuzzy Systems*. – 2006. – Vol.14. – № 3. – P. 468 – 477.

2. Демидова, Л.А. Прогнозирование тенденций рынка труда на основе однофакторных нечетких временных рядов // Системы управления и информационные технологии. – 2007. – №3,2(2)9. – С. 241 – 246.