ПРОГНОЗИРОВАНИЕ ТЕНДЕНЦИЙ ВРЕМЕННЫХ РЯДОВ НА ОСНОВЕ ОДНОФАКТОРНОЙ НЕЧЕТКОЙ МОДЕЛИ С ИСПОЛЬЗОВАНИЕМ ДИСКРЕТНЫХ НЕЧЕТКИХ МНОЖЕСТВ ВТОРОГО ТИПА И ГЕНЕТИЧЕСКОГО АЛГОРИТМА

Л.А. Демидова,
к.т.н., доцент кафедры вычислительной и прикладной математики Рязанского государственного радиотехнического университета,
e-mail Liliya.Demidova@rambler.ru

Рассматривается применение дискретных нечетких множеств второго типа для разработки однофакторных нечетких моделей прогнозирования. Предлагается генетический алгоритм, обеспечивающий выбор оптимальных параметров модели прогнозирования – действительных чисел для корректировки границ универсума, числа интервалов разбиения универсума и степеней принадлежности элементов дискретных нечетких множеств второго типа.

Введение

Анализ временных рядов играет важную роль в решении многих актуальных задач, например, при краткосрочном прогнозировании тенденций рынка труда в России. В настоящее время существует необходимость в разработке методов прогнозирования, которые бы обеспечили получение адекватной оценки предстоящих изменений политики и принятия решений в региональных органах управления на основе известных показателей развития регионов. Так как большинство реальных событий характеризуются некоторой неопределенностью, то каждому наблюдению временного ряда (фактора) можно поставить в соответствие нечеткую переменную с некоторой функцией принадлежности.

Модель прогнозирования на основе нечетких множеств первого типа

Нечеткие временные ряды могут быть представлены с помощью нечетких множеств первого или второго типа [1, 2, 3].

Дискретное нечеткое множество первого типа (ДНМТ1) A, определенное на универсуме может быть определено в виде:

$$\mathrm{A} = f_A(u_1)/u_1 + f_A(u_2)/u_2 + \ldots + f_A(u_n)/u_n\,, \quad (1)$$

где $f_A(u)$ – функция принадлежности ДНМТ1 A, $f_A(u)\colon U \to [0,\,1]$, $f_A(u_r)$ определяет степень принадлежности элемента u_r ДНМТ1 A, $r = \overline{1,n}$.

Пусть $Y(t)$ ($t = \ldots, 0, 1, 2, \ldots$) – универсум, определенный на множестве действительных чисел, а $F(t)$ – набор функций $f_i(t)$ ($i = 1, 2, \ldots$)), определенных на универсуме $Y(t)$. Тогда $F(t)$ называется нечетким временным рядом на универсуме $Y(t)$.

Пусть

$$F(t) = F(t-1) \circ R(t, t-1),$$

где $R(t, t-1)$ – нечеткое отношение и $\circ$ – операция *max-min* композиции.

Обозначим зависимость $F(t)$ от $F(t-1)$ как $F(t-1) \to F(t)$, где $F(t-1)$ и $F(t)$ – нечеткие множества.

Если $F(t)$ зависит от $F(t–1)$, $F(t–2)$,..., $F(t–k)$, то нечеткая логическая зависимость представляется как: $F(t–k)$,..., $F(t–2)$, $F(t–1)\rightarrow F(t)$ и называется однофакторной k-порядковой моделью прогнозирования на основе нечетких временных рядов.

Рассмотрим модель прогнозирования при $k = 1$.

$$F(t-1)\rightarrow F(t). \qquad (2)$$

Представим нечеткие данные i-го и $(i+1)$-го периодов (дней, месяцев, кварталов и т.п.) как нечеткие множества A_j и A_k на универсуме U. Тогда нечеткая логическая зависимость может быть представлена в виде: $A_j \rightarrow A_k$, где A_j – текущее состояние, а A_k – следующее состояние нечеткой зависимости.

Пусть $f_i(t) = T$, $(i = 1, 2, ...)$ – 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 $D_{\min}$ и $D_{\max}$ – минимальное и максимальное приращения значений фактора на основе известных данных соответственно ($D_{\min} = \min(f(t)–f(t–1)$, $D_{\max} = \max(f(t)–f(t–1))$, а D_1 и D_2 – два действительных числа, обеспечивающие разбиение универсума U на интервалы равной длины: $u_1, u_2, ..., u_n$ [1, 2]. Лингвистические термы A_r $(r = \overline{1, n})$, представленные нечеткими множествами фактора, имеют вид:

$$A_1 = 1/u_1 + 0{,}5/u_2 + 0/u_3 + ... + 0/u_{n-1} + 0/u_n\,,$$
$$A_2 = 0{,}5/u_1 + 1/u_2 + 0{,}5/u_3 + 0/u_4 + ... + 0/u_n\,,$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$A_n = 0/u_1 + 0/u_2 + ... + 0/u_{n-2} + 0{,}5/u_{n-1} + 1/u_n\,.$$

Если значение приращения фактора принадлежит интервалу u_1, то соответствующее нечеткое значение имеет вид: $X_n = 1/A_1 + 0{,}5A_2$. Если значение приращения фактора принадлежит интервалу u_r, то соответствующее нечеткое значение имеет вид: $X_r = 0{,}5/A_{r-1} + 1A_r + 0{,}5/A_{r+1}$, $r = \overline{2, n-1}$. Если значение приращения фактора принадлежит интервалу u_n, то нечеткое значение имеет вид: $X_n = 0{,}5/A_{n-1} + 1A_n$.

Пусть X_k и X_j – нечеткие значения приращения для i-го и $(i + 1)$-го периодов соответственно. Для i-го периода можно записать нечеткую логическую зависимость вида: $X_k \rightarrow X_j$. На основе нечетких зависимостей для всех известных значений временного ряда определяются группы зависимостей путем объединения зависимостей с одинаковой левой частью в одну группу. Так, зависимости $X_k \rightarrow X_j$, $X_k \rightarrow X_l$, $X_k \rightarrow X_s$ объединяются в группу: $X_k \rightarrow X_j, X_k\, X_s$, а функция принадлежности ДНМТ1 группы определяется как $f_X(u_r) = \max\limits_{r=1,\ n}(f_{X_j}(u_r), f_{X_l}(u_r), f_{X_s}(u_r))$ [2].

Результирующее ДНМТ1 для прогнозируемого значения временного ряда для $(i+1)$-го периода находится как объединение ДНМТ1, входящих в правую часть группы нечетких зависимостей для i-го периода [1, 2].

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}:

$$F_{i+1} = T_i + y_{i+1}\,. \qquad (3)$$

Четкое значение приращения фактора для $(i+1)$-го периода находится по методу центра тяжести для одноточечных множеств

$$y_{i+1} = \sum_{r=1}^{n} w_r \cdot z_r \Big/ \sum_{r=1}^{n} w_r\,, \qquad (4)$$

где n – количество интервалов u_r $(r = (r = \overline{1, n})$;

z_r – средняя точка r-го интервала;

w_r – значение степени принадлежности для r-го интервала результирующего ДНМТ1, описывающего группу нечетких зависимостей.

Средняя относительная ошибка прогноза (AFER – average forecasting error rate) может быть вычислена по формуле:

$$AFER = (100/m)\cdot \sum_{i=1}^{m} \left|(F_i - T_i)/T_i\right|, \qquad (5)$$

где F_i и T_i – предсказанное и реальное значения для i-го периода;

m – количество значений (периодов) временного ряда.

Модель прогнозирования на основе нечетких множеств второго типа

Нечеткие множества второго типа позволяют моделировать различные неопределенности, которые не могут быть адекватно представлены с помощью нечетких множеств первого типа [3]. Однако применение нечетких множеств второго типа обычно увеличивает вычислительную сложность по сравнению с нечеткими множествами первого типа из-за наличия дополнительной размерности. Поэтому, использование нечетких множеств второго типа является целесообразным, если позволяет обеспечить значительное улучшение результатов (например, повышение точности прогноза).

Интервальное дискретное нечеткое множество второго типа (ИДНМТ2) $\tilde{A}$, определенное на универсуме U может быть определено в виде [3]:

$$\tilde{A} = f_{\tilde{A}}(u_1)/u_1 + f_{\tilde{A}}(u_2)/u_2 + \ldots + f_{\tilde{A}}(u_n)/u_n, \quad (6)$$

где $f_{\tilde{A}}(u) = \underline{\mu}_{\tilde{A}}(u), \bar{\mu}_{\tilde{A}}(u)$; $\underline{\mu}_{\tilde{A}}(u), \bar{\mu}_{\tilde{A}}(u)$ – «нижняя» и «верхняя» функции принадлежности ИДНМТ2, являющиеся функциями принадлежности ДНМТ1, характеризующие «отпечаток неопределенности» (footprint of uncertainty) *FOU* [3]; $f_{\tilde{A}}(u){:}\,U \to [0, 1], f_{\tilde{A}}(u_r)\ (r = \overline{1,n})$ – степень принадлежности элемента u_r по «нижней» и «верхней» функциям принадлежности ИДНМТ2.

На *рис. 1* приведен пример *FOU* для ИДНМТ2. Серым цветом закрашен сам *FOU*, символы ○ и ● определяют дискретные значения для «нижней» и «верхней» функций принадлежности ИДНМТ2 соответственно (при совпадении значений для «нижней» и «верхней» функций принадлежности символ ● изображен внутри символа ○).

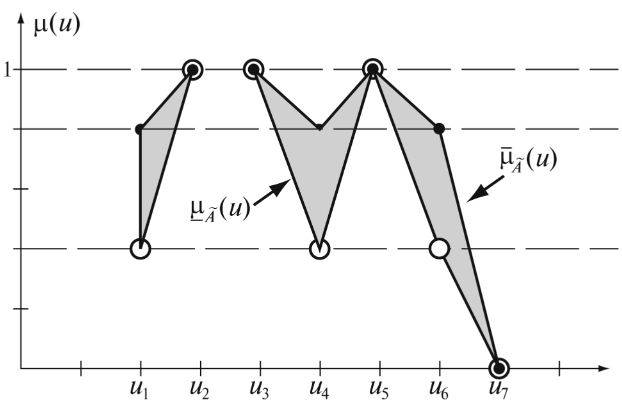

Рис. 1. FOU для ИДНМТ2

Лингвистические термы $\tilde{A}$ $(r = \overline{1,n})$ на основе ИДНМТ2 могут быть определены в виде:

$$\tilde{A}_1 = 1/u_1 + V/u_2 + 0/u_3 + \ldots + 0/u_{n-1} + 0/u_n,$$
$$\tilde{A}_2 = V/u_1 + 1/u_2 + V/u_3 + 0/u_4 + \ldots + 0/u_n,$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$\tilde{A}_n = 0/u_1 + 0/u_2 + \ldots + 0/u_{n-2} + 0{,}5/u_{n-1} + 1/u_n,$$

где $V = \alpha_{lower}, \alpha_{upper}$; α_{lower} и α_{upper} – значения «нижней» функций принадлежности $\underline{\mu}_{\tilde{A}}(u)$ и «верхней» функций принадлежности $\bar{\mu}_{\tilde{A}}(u)$ в точке u_r $(r = \overline{1,n})$.

Таким образом, каждому лингвистическому терму $\tilde{A}_r$ соответствует некоторый *FOU*, границы которого определяются с помощью «нижней» и «верхней» функций принадлежности $\underline{\mu}_{\tilde{A}}(u), \bar{\mu}_{\tilde{A}}(u)$.

Модель прогнозирования на основе ИДНМТ2 строится аналогично модели на основе ДНМТ1. Пусть определены FOU_k и FOU_j для приращений фактора для *i*-го и (*i*+1)-периодов соответственно. Тогда для *i*-го периода можно записать нечеткую логическую зависимость вида: $FOU_k \to FOU_j$. Группы нечетких зависимостей определяются путем объединения нечетких зависимостей с одинаковой левой частью в одну группу. Если были сформированы зависимости: $FOU_k \to FOU_j$, $FOU_k \to FOU_l$, $FOU_k \to FOU_s$, то они объединяются в группу: $FOU_k \to FOU_j$, FOU_k, FOU_l, FOU_s, а «нижняя» и «верхняя» функции принадлежности ДНМТ1, характеризующие группу, определяются как

$$f_{\tilde{A}}(u_r) = \max_{r=\overline{1,n}}(f_{\tilde{A}_j}(u_r), f_{\tilde{A}_l}(u_r), f_{\tilde{A}_s}(u_r)), \quad V = \alpha_{lower}, \alpha_{upper}.$$

Результирующие ДНМТ1 для прогнозируемого значения (*i*+1)-го периода находятся как объединения ДНМТ1, соответствующих «нижней» и «верхней» функциям принадлежности, для *FOU*, входящих в правую часть группы нечетких зависимостей для *i*-го периода. Таким образом, для каждого прогнозируемого значения (*i*+1)-го периода временного ряда находится соответствующие ему ИДНМТ2 и *FOU* – выходной «отпечаток неопределенности».

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} по формуле (3).

Четкое значение приращения фактора (центроид) для (*i*+1)-го периода находится с помощью операции «type-reduction» (операции «понижения типа») [3]. При этом определяют два «вложенных» нечетких множества первого типа – *L* и *R* – внутри *FOU* интервального нечеткого множества второго типа $\tilde{A}$. Множества *L* и *R* имеют минимально и максимально возможный центроиды в $\tilde{A}$ соответственно. Результирующее четкое значение центроида определяется как среднее значение от центроидов множеств *L* и *R*.

ИДНМТ2 может быть представлено интервалом определения, который описывается с помощью его левой и правой конечных точек как $[y_{left}, y_{right}]$ соответствующих множествам *L* и *R*, или с помощью его центра и протяженности как $[c–s, c + s]$, где $c = (y_{left} + y_{right})/2$, $s = (y_{left} – y_{right})/2$ [3].

Центроид ДНМТ1 представляет собой взвешенное среднее по формуле (4). Для вычисления центроида ИДНМТ2 $C_{\tilde{A}}$ необходимо представить w_r (значение степени принадлежности для *r*-го интервала результирующего ДНМТ1 в формуле (4)) как ДНМТ1. Центроид $C_{\tilde{A}}$ определяется через центроиды всех «вложенных» в *FOU* ДНМТ1 и может быть представлен как:

$$C_{\tilde{A}} = \int_{z_1 \in Z_1} \ldots \int_{z_n \in Z_n} \int_{w_1 \in W_1} \ldots \int_{w_n \in W_n} 1/(\sum_{r=1}^{n} w_r \cdot z_r / \sum_{r=1}^{n} w_r) = [y_{left},\ y_{right}], \quad (7)$$

где Z_r ($r=\overline{1,n}$) в (7) представляет собой ДНМТ1, имеющее центр c_r и протяженность s_r ($s_r \geq 0$), а W_r ($r=\overline{1,n}$) представляет собой ДНМТ1, имеющее центр h_r и протяженность Δ_r ($\Delta_r \geq 0$).

Так как Z_1, ..., Z_n, W_1, ..., W_n представляют собой ДНМТ1, то и $C_{\tilde{A}}$ является ДНМТ1. Для вычисления центроида $C_{\tilde{A}}$ необходимо найти две конечные точки интервала: y_{left} и y_{right} и , рассмотрев задачи минимизации и максимизации функции (положив $z_r = c_r + s_r$ и $z_r = c_r - s_r$ соответственно) [3]:

$$y(w_1,...,w_n) = \sum_{r=1}^{n} w_r \cdot z_r / \sum_{r=1}^{n} w_r \quad (8)$$

при условии $w_r \in [h_r - \Delta_r,\ h_r + \Delta_r]$, $h_r \geq \Delta_r$, $r=\overline{1,n}$.

Продифференцируем функцию $y(w_1,...,w_n)$ по w_k:

$$\frac{\partial}{\partial w_k} y(w_1,...,w_n) = \frac{\partial}{\partial w_k} (\sum_{r=1}^{n} w_r \cdot z_r / \sum_{r=1}^{n} w_r) =$$
$$= (z_k - y(w_1,...,w_n)) / \sum_{r=1}^{n} w_r. \quad (9)$$

Так как $\sum_{r=1}^{n} w_r > 0,$

то из (9) следует, что:

$$\frac{\partial}{\partial w_k} y(w_1,...,w_n) \begin{matrix}\geq\\=\\\leq\end{matrix} 0, \text{ если } w_k \begin{matrix}\geq\\=\\\leq\end{matrix} y(w_1,...,w_n) \quad (10)$$

Так как из $y(w_1,...,w_n) = z_k$ следует, что

$$\sum_{r=1}^{n} w_r \cdot z_r / \sum_{r=1}^{n} w_r = z_k, \text{ то}$$

$$\sum_{r=1, r\neq k}^{n} w_r \cdot z_r / \sum_{\neq 1, r\neq k}^{n} w_r = z_k. \quad (11)$$

Из (10) видно, что если $z_k > y(w_1,...,w_n)$, то $y(w_1,...,w_n)$ увеличивается при увеличении w_k; а если $z_k < y(w_1,...,w_n)$, то $y(w_1,...,w_n)$ уменьшается при уменьшении w_k. Для вычисления центроида $C_{\tilde{A}}$ можно использовать итерационный алгоритм Карника-Менделя [3].

Итерационный алгоритм Карника-Менделя

Пусть $h_r \geq \Delta_r$ так, что $w_r \geq 0$ для $r=\overline{1,n}$. Максимальное (минимальное) значение, которое может принимать w_k ($k=\overline{1,n}$), равно $h_k + \Delta_k$ ($h_k - \Delta_k$).

Функция $y(w_1,...,w_n)$ достигает своего максимального значения, если:

- ✧ $w_k = h_k + \Delta_k$ для тех значений k, для которых $z_k > y(w_1,...,w_n)$;
- ✧ $w_k = h_k - \Delta_k$ для тех значений k, для которых $z_k < y(w_1,...,w_n)$.

Функция $y(w_1,...,w_n)$ достигает своего минимального значения, если:

- ✧ $w_k = h_k - \Delta_k$ для тех значений k, для которых $z_k > y(w_1,...,w_n)$;
- ✧ $w_k = h_k + \Delta_k$ для тех значений k, для которых $z_k < y(w_1,...,w_n)$.

Максимум функции $y(w_1,...,w_n)$ может быть определен с помощью следующей итерационной процедуры. Пусть $z_r = c_r + s_r$, ($r=\overline{1,n}$). Предположим, что все z_r упорядочены по возрастанию, то есть $z_1 \leq z_2 \leq ... \leq z_n$.

1. Пусть $w_r = h_r$ для $r=\overline{1,n}$. Вычислим $y' = y(h_1,...,h_n)$ по формуле (8).
2. Определим такое $k(1 \leq k \leq n-1)$, что $z_k \leq y' \leq z_{k+1}$.
3. Пусть $w_r = h_r - \Delta_r$ для $r \leq k$ и $w_r = h_r + \Delta_r$ для $r \geq k+1$. Вычислим $y'' = y(h_1 - \Delta_1,...,h_k - \Delta_k,\ h_{k+1} + \Delta_{k+1},...,h_n + \Delta_n)$ по формуле (8). Так как все z_r упорядочены по возрастанию, с учетом формул (10) и (11) и неравенства $z_k \leq y' \leq z_{k+1}$, можно утверждать, что $y'' \geq y'$, поскольку w_r выбраны так, что w_r уменьшены для $r \leq k$ и w_r увеличены для $r \geq k+1$.
4. Если $y' = y''$, то вычисления заканчиваются, а y'' представляет собой максимум функции $y(w_1,...,w_n)$. Если $y' \neq y''$, то осуществляется переход к шагу 5.
5. Полагаем $y' = y''$ и осуществляем переход к шагу 2.

Алгоритм требует не более n итераций, где одна итерация состоит из шагов 2–5 [3].

Минимум функции $y(w_1,...,w_n)$ может быть определен с помощью аналогичной итерационной процедуры, где $z_r = c_r + s_r$, ($r=\overline{1,n}$); а на Шаге 3 для вычисления $y'' = y(h_1+\Delta_1,...,h_k+\Delta_k,\ h_{k+1}-\Delta_{k+1},...,h_n+\Delta_n)$ полагается, что $w_r + \Delta_r$ для $r \leq k$ и $w_r = h_r - \Delta_r$ для $r \geq k+1$.

Генетический алгоритм, обеспечивающий повышение точности прогнозирования на основе нечетких множеств второго типа

Самостоятельной задачей при прогнозировании на основе нечетких временных рядов является определение оптимальных параметров модели, обеспечивающих максимальную точность прогнозирования: действительных чисел D_1, D_2, используемых при корректировке универсума U, количества интервалов разбиения $D_3 = n$ универсума U и степеней принадлежности элементов D_4 (α_{upper}), D_5 (α_{lower}) для ИДНМТ2. Применение генетического алгоритма (ГА) позволяет значительно сократить время поиска оптимальных значений параметров D_1, D_2, D_3, D_4, D_5 [5]. При этом хромосома s будет иметь вид: $s = (D_1, D_2, D_3, D_4, D_5)$. Для каждого элемента хромосомы следует задать диапазоны их изменения: для D_1 – $[-d_1; 0]$, для D_2 – $[0; d_2]$, для D_3 = n–$[2; n_{\max}]$, для D_4, D_5 – $[0; 1]$, где d_1, d_2, – положительные действительные числа, равные, например, $d_i = D_{\max} - D_{\min}$, $i = 1,2$; $n_{\max}$ – натуральное число, $n_{\max} \le m-1$ (m – количество значений временного ряда). Также при формировании начальной популяции, при выполнении операций скрещивания и мутации необходимо следить за выполнением требования: $D_4 \ge D_5$ так как элементы D_4, D_5,, в хромосоме определяют верхнее и нижнее значение функций принадлежности ИДНМТ2 соответственно.

Для каждого набора параметров D_1, D_2, D_3, D_4, и D_1, D_2, D_3, D_5 необходимо вычислить функцию соответствия. В качестве функции соответствия можно выбрать функцию (5). Однако при вычислении функции соответствия по формуле (5) для хромосом как начальной популяции размера P, так и расширенной популяции размера $(P + P^*P_c)$ (где P_c – коэффициент скрещивания) может быть получено значение вида «0/0» (если имеются группы нечетких зависимостей с неопределенными правыми частями, и поэтому невозможно вычислить прогнозное значение по формулам (3), (4), так как значение приращения фактора y_{i+1} определяется как «0/0»). Поэтому при удалении из популяции размером $(P + P^*P_c)$ хромосом с худшими значениями функций соответствия следует предварительно оценить хромосомы со значением функции соответствия «0/0» как наихудшие (например, таким хромосомам можно поставить в соответствие максимально возможное значение ошибки, равное 100%). При сортировке хромосом по возрастанию значений функций соответствия хромосомы со значением функции соответствия, равным «0/0», будут занимать последние места в списке и в результате отбраковки P^*P_c худших хромосом будут исключаться из популяции [1]. Для обеспечения гарантированного выполнения прогноза следующего значения временного ряда необходимо видоизменить функцию соответствия в ГА следующим образом. Если для каждого набора D_1, D_2, D_3, D_4, и D_1, D_2, D_3, D_5 некоторой хромосомы s определены все правые части групп нечетких зависимостей, то функция соответствия для этих наборов вычисляется по формуле (5). Если для любого из наборов D_1, D_2, D_3, D_4, и D_1, D_2, D_3, D_5 не определена хотя бы одна правая часть в группах нечетких зависимостей, то значение функции соответствия находится как сумма средней относительной ошибки прогноза по формуле (5) и числа 100.

Таким образом, видоизмененная функция соответствия имеет вид:

$$J_V(s) = \begin{cases} AFER, & \text{если определены все правые части в группах логических зависимостей} \\ AFER + 100, & \text{если не определена хотя бы одна правая часть в группах логических зависимостей} \end{cases}, \qquad (12)$$

где $AFER$ определяется по формуле (5), $V = \alpha_{lower}$, α_{upper}.

В результате, набор D_1, D_2, D_3, D_4 или D_1, D_2, D_3, D_5, несмотря на то, что для него средняя относительная ошибка прогноза по формуле (5) может быть минимальной (если значение функции соответствия не определяется как «0/0»), будет признан одним из худших в процессе реализации ГА, и, возможно, соответствующая ему хромосома будет исключена из популяции.

На основе наборов D_1, D_2, D_3, D_4 и D_1, D_2, D_3, D_5, вычисляются два значения функций соответствия $J_{\alpha_{upper}}$ и $J_{\alpha_{lower}}$. Если хотя бы для одного из наборов значение функции соответствия оказалось больше 100, то соответствующую ему хромосому следует признать «нежизнеспособной» и положить значение ее функции соответствия J_s равным наибольшему из двух значений функций соответствия наборов D_1, D_2, D_3, D_4 и D_1, D_2, D_3, D_5. При этом нет необходимости в вычислении функции соответствия для хромосомы с использованием алгоритма Карника-Менделя для определения центроида ИДНМТ2. В противном случае для хромосомы s вычисляется средняя относительная ошибка прогнозирования $AFER$ по формуле (5) с использованием алгоритма Карника-Менделя для определения центроида ИДНМТ2. Если значение $AFER$ для s (и ИДНМТ2) окажется меньше, чем значения функций соответствия $J_{\alpha_{upper}}$ и $J_{\alpha_{lower}}$ и для наборов D_1, D_2, D_3, D_4 и D_1, D_2, D_3, D_5 соответственно, то такую хромосому следует считать «жизнеспособной», а значение ее функции соответствия J_s положить равным $AFER$, иначе необходимо положить значение ее функции соответствия J_s равным $AFER + 100$ (для возможного исключения этой хромосомы из популяции). Таким образом, функцию соответствия хромосомы s следует вычислять по формуле:

$$J_s(s) = \begin{cases} AFER, & \text{при } AFER < J_{\alpha_{upper}},\ AFER < J_{\alpha_{lower}} \\ AFER + 100, & \text{иначе} \end{cases}. \qquad (13)$$

Отметим, что необходимо затратить дополнительное время на формирование начальной популяции, чтобы она состояла только из «жизнеспособных» хромосом (у которых значение функции соответствия меньше 100). Тогда выполнение операций скрещивания и мутации будет более эффективным и результативным (иначе вся популяция может быть с самого начала «нежизнеспособной»). Хромосома, минимизирующая функцию (13), имеет больше шансов быть признанной лучшей. Выбор родителя будет состоять в выборе лучшей хромосомы из двух случайно выбранных. Затем две выбранные таким образом хромосомы-родителя используются для скрещивания, при этом выбирается коэффициент скрещивания R_c и генерируется число $N_c = rahdom([0, 1)]$. Если $R_c > N_c$, то случайным образом выбирается точка скрещивания z и выполняется скрещивание. При выполнении мутации выбирается коэффициент мутации R_m и генерируется число $N_m = rahdom([0, 1)]$. Если $R_m > N_m$, то случайным образом выбирается точка мутации z.

Тогда генетический алгоритм имеет вид:

1. Создается начальная популяция размера P из случайным образом выбранных хромосом s.

2. При $g < G$ (G – количество генераций) вычисляется функция соответствия для каждой хромосомы, затем создается $P/2$ пар хромосом-родителей и осуществляется переход к шагу 3. При $g > G$ осуществляется переход к шагу 5.

3. Выполняются операции скрещивания и мутации для текущей популяции.

4. Создается новая популяция размера P, дополненная хромосомами-детьми, а хромосомы с худшими значениями функции соответствия отбрасываются.

5. Выбирается лучшая хромосома, которая минимизирует функцию соответствия.

Пример прогнозирования

На примере данных по фактору «численность занятого населения» (в России) для периодов 2-1999 – 5-2004, полученных от Госкомстата (*табл. 1*) была построена нечеткая модель прогнозирования на основе ИДНМТ2. Представление выходных значений фактора на основе ИДНМТ2 приведено в *табл. 1.* При этом были получены группы нечетких логических зависимостей:

Группа 1: $FOU_1 \rightarrow FOU_2$.
Группа 2: $FOU_2 \rightarrow FOU_5$, FOU_6.
Группа 3: $FOU_3 \rightarrow FOU_2$, FOU_3, FOU_4 FOU_7.
Группа 4: $FOU_4 \rightarrow FOU_3$.
Группа 5: $FOU_5 \rightarrow FOU_1$, FOU_3, FOU_5.
Группа 6: $FOU_6 \rightarrow FOU_5$.
Группа 7: $FOU_7 \rightarrow FOU_4$.

Рис. 2. Графические зависимости для реальных и прогнозируемых значений

В *табл. 2* приведены параметры трех однофакторных нечетких моделей и результаты прогнозирования. «Модель 1» основана на ДНМТ1 при заранее заданном значении степени принадлежности элементов нечеткому множеству . Параметры определялись с помощью ГА. «Модель 2» основана на ДНМТ1. Параметры определялись с помощью ГА. «Модель 3» основана на ИДНМТ2. Параметры определялись с помощью ГА. На рис. 2 приведены графические зависимости для реальных и прогнозируемых значений на основе ИННМТ2. Анализ результатов моделирования показывает уменьшение средней относительной ошибки прогнозирования за счет применения ИДНМТ2 (при этом относительная ошибка прогнозирования для периода 5-2004 на основе «Модели 3» равна 1,173505 %).

Заключение

Предлагаемый метод прогнозирования тенденций рынка труда обеспечивает получение более высоких результатов прогноза, чем предложенный в [4], и может быть рекомендован для проведения краткосрочных прогнозов. Применение ИДНМТ2 и ГА для поиска оптимальных параметров нечеткой модели обеспечили более высокую точность прогноза. Для повышения точности прогноза можно представлять значения временных рядов с помощью непрерывных нечетких множеств второго типа, что приводит, однако, к соответствующему увеличению вычислительной сложности и временных затрат. ■

Таблица 1

Фактор «численность занятого населения»

Месяц-год	Фактор, тыс. ч.	Приращение, тыс. ч.	Входное ИДНМТ2	Выходное ИДНМТ2 ($V = \alpha_{lover}, \alpha_{upper}$)
2-99	60614	–	–	–
5-99	62462	1848	FOU_6	–
8-99	63742	1280	FOU_5	$0/A_1+0/A_2+0/A_3+V/A_4+1/A_5+V/A_6+0/A_7$
11-99	63082	-660	FOU_3	$1/A_1+V/A_2+1/A_3+V/A_4+1/A_5+V/A_6+0/A_7$
2-00	62439	-643	FOU_3	$V/A_1+1/A_2+1/A_3+1/A_4+V/A_5+V/A_6+1/A_7$
5-00	64961	2522	FOU_7	$V/A_1+1/A_2+1/A_3+1/A_4+V/A_5+V/A_6+1/A_7$
8-00	65154	193	FOU_4	$0/A_1+0/A_2+V/A_3+1/A_4+V/A_5+0/A_6+0/A_7$
11-00	64465	-689	FOU_3	$0/A_1+V/A_2+1/A_3+V/A_4+1/A_5+V/A_6+0/A_7$
2-01	62953	-1512	FOU_2	$V/A_1+1/A_2+1/A_3+1/A_4+V/A_5+V/A_6+1/A_7$
5-01	64542	1589	FOU_6	$0/A_1+0/A_2+0/A_3+V/A_4+1/A_5+1/A_6+V/A_7$
8-01	65459	917	FOU_5	$0/A_1+0/A_2+0/A_3+V/A_4+1/A_5+V/A_6+0/A_7$
11-01	64664	-795	FOU_3	$1/A_1+V/A_2+1/A_3+V/A_4+1/A_5+V/A_6+0/A_7$
2-02	65021	357	FOU_4	$V/A_1+1/A_2+1/A_3+1/A_4+V/A_5+V/A_6+1/A_7$
5-02	65962	941	FOU_5	$0/A_1+V/A_2+1/A_3+V/A_4+1/A_5+V/A_6+0/A_7$
8-02	67502	1540	FOU_5	$1/A_1+V/A_2+1/A_3+V/A_4+1/A_5+V/A_6+0/A_7$
11-02	65766	-1736	FOU_1	$1/A_1+V/A_2+1/A_3+V/A_4+1/A_5+V/A_6+0/A_7$
2-03	64104	-1662	FOU_2	$V/A_1+1/A_2+V/A_3+0/A_4+0/A_5+0/A_6+0/A_7$
5-03	65528	1424	FOU_5	$0/A_1+0/A_2+0/A_3+V/A_4+1/A_5+1/A_6+V/A_7$
8-03	66674	1146	FOU_5	$1/A_1+V/A_2+1/A_3+V/A_4+1/A_5+V/A_6+0/A_7$
11-03	66496	-178	FOU_3	$1/A_1+V/A_2+1/A_3+V/A_4+1/A_5+V/A_6+0/A_7$
2-04	64941	-1555	FOU_2	$V/A_1+1/A_2+1/A_3+1/A_4+V/A_5+V/A_6+1/A_7$
5-04	67271	2330	FOU_6	$0/A_1+0/A_2+0/A_3+V/A_4+1/A_5+1/A_6+V/A_7$

Таблица 2

Параметры нечётких моделей

Параметры	«Модель 1»	«Модель 2»	«Модель 3»
D_1	816,486940898299	818,938883168293	818,914669508277
D_2	662,918661869601	656,198590769605	656,765458625010
$D_3 = n$	7	7	7
$D_0 = \alpha$	0,5	0	–
α_{upper} (α_{lower})	–	–	1 (0)
AFER (%)	1,22965304295085	1,22676137780468	1,22528803913897
$AFER_{upper}$ (%)	–	–	1,24229944946879
$AFER_{lower}$ (%)	–	–	1,22678992137390
t (с)	36,782	91,203	187,734

Литература

1. Демидова Л.А. Прогнозирование тенденций рынка труда на основе однофакторных нечетких временных рядов // Системы управления и информационные технологии, 2007, № 3.2(29). – С. 241–246.
2. Chen S.M. Forecasting enrollments based on fuzzy time series // Fuzzy Sets Systems, 1996, vol. 81, no. 3, pp. 311–319.
3. Mendel J. M. Type-2 fuzzy sets and systems: an overview // IEEE Computational intellegence magazine. 2007, vol. 2, № 1, pp. 20–29.
4. Ярушкина Н.Г. Основы теории нечетких и гибридных систем: Учеб. пособие. – М.: Финансы и статистика, 2004, 320 с.: ил.