Демидова Л.А., Соколов А.А., Токмаков Ю.А. Прогнозирование экономических показателей на основе нечетких множеств и генетического алгоритма. // Проблемы информатики в образовании, управлении, экономике и технике: Сб. статей VIII Всерос. научно-техн. конф. – Пенза: ПДЗ, 2008. – С. 256-263.

ПРОГНОЗИРОВАНИЕ ЭКОНОМИЧЕСКИХ ПОКАЗАТЕЛЕЙ НА ОСНОВЕ НЕЧЕТКИХ МНОЖЕСТВ И ГЕНЕТИЧЕСКОГО АЛГОРИТМА

Л.А. Демидова, А.А. Соколов, Ю.А. Токмаков

Рязанский государственный радиотехнический университет,
г. Рязань

В настоящее время существует необходимость в разработке методов прогнозирования, которые бы обеспечили получение адекватной оценки предстоящих изменений в экономике на основе известных показателей. Анализ временных рядов играет важную роль в решении многих актуальных задач. Так как большинство реальных событий характеризуются некоторой неопределенностью, то каждому наблюдению нечеткого временного ряда (фактору) можно поставить в соответствие нечеткую переменную с некоторой функцией принадлежности [2]. С помощью нечетких отношений и нечетких правил можно построить эффективную модель краткосрочного прогнозирования на основе большого числа входов и одного выхода (без значительных временных затрат и вычислительных сложностей).

Нечеткие временные ряды могут быть представлены с помощью нечетких множеств [1, 3]. Нечеткое множество (НМ) A, определенное на универсуме U может быть определено в виде:

$$A = f_A(u_1)/u_1 + f_A(u_2)/u_2 + \ldots + f_A(u_n)/u_n ,$$

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

Пусть $Y(t)$ ($t = \ldots,\ 0,\ 1,\ 2,\ldots$) – некоторый универсум, определенный на множестве действительных чисел. Предположим, что $F(t)$ представляет собой набор функций $f_i(t)$ ($i = 1,\ 2,\ldots$), определенных на универсуме $Y(t)$ ($t = \ldots,\ 0,\ 1,\ 2,\ldots$). 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 $max\text{-}min$-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 $F(t-1)$, $F(t-2), \ldots,$ $F(t-k)$, то нечеткая логическая зависимость представляется как $F(t-k), \ldots, F(t-2)$, $F(t-1) \to F(t)$ и называется однофакторной k – порядковой моделью прогноза на основе нечетких временных рядов.

Если помимо фактора X используется еще и фактор Y, тогда модель называется двухфакторной. Фактор X является главным, фактор Y – вспомогательным. Если имеется m факторов, то модель называется m-факторной.

Пусть $Y(t)$ ($t = \dots,\ 0,\ 1,\ 2,\dots$) – универсум и $Y(t) \subseteq R$. Предположим, что $f_i(t)$ ($i = 1,\ 2,\dots$) определены на универсуме $Y(t)$, и предположим, что $F(t)$ представляет собой набор (коллекцию) $f_i(t)$ ($i = 1,\ 2,\dots$). Тогда $F(t)$ называется нечеткой временной серией на $Y(t)$ ($t = \dots,\ 0,\ 1,\ 2,\dots$).

Пусть $F(t)$ – нечеткая временная серия и $F(t) = F(t-1) \circ R(t,t-1)$, где $R(t,t-1)$ – нечеткое отношение и $\circ$ – операция $max-min$-композиции. Тогда $F(t)$ является следствием от $F(t-1)$, и это обозначается как $F(t-1) \to F(t)$, где $F(t-1)$ и $F(t)$ – нечеткие множества.

С целью прогноза определим нечеткое отношение между текущим и будущим состояниями временных рядов с помощью нечетких множеств. Пусть фаззифицированные данные i-го и $i+1$-го дней представлены через нечеткие множества A_j и A_k, определенные на универсуме U. Тогда нечеткое логическое отношение может быть представлено через $A_j \to A_k$, где A_j – текущее состояние нечеткого логического отношения, а A_k – следующее состояние нечеткого логического отношения.

Используя нечеткие композиционные правила, получим систему нечеткого вывода для прогнозирования нечетких временных рядов с высокой точностью. Точность прогноза может быть увеличена за счет использования большего числа факторов и большей зависимости от исторических данных.

Рассмотрим алгоритм прогнозирования с помощью m-факторной k – порядковой модели прогноза на основе нечетких временных серий.

Шаг 1. Определим универсум U для главного фактора как $U = [D_{min} - D_1,\ D_{max} - D_2]$, где D_{min} и D_{max} – минимальное и максимальное значения главного фактора на основе известных (исторических) данных соответственно, а D_1 и D_2 – два таких действительных числа, использование которых позволяет разбить универсум U на n подинтервалов равной длины: u_1, u_2, …, u_n.

Определим универсумы V_i ($i = \overline{1, m-1}$) для вспомогательных факторов аналогично определению универсума для главного фактора.

Шаг 2. Определим лингвистические термы A_r ($r = \overline{1,\ n}$), описывающие значения главного фактора и представленные нечеткими множествами в виде:

$$A_1 = 1/u_1 + 0{,}5/u_2 + 0/u_3 + \dots + 0/u_{n-1} + 0/u_n\ ,$$

$$A_2 = 0{,}5/u_1 + 1/u_2 + 0{,}5/u_3 + 0/u_4 + \dots + 0/u_n\ ,$$

$$A_3 = 0/u_1 + 0{,}5/u_2 + 1/u_3 + 0{,}5/u_4 + \dots + 0/u_n\ ,$$

$$\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots ,$$

$$A_n = 0/u_1 + 0/u_2 + \dots + 0/u_{n-2} + 0{,}5/u_{n-1} + 1/u_n\ .$$

Аналогично для i-го нечеткого временного ряда определим лингвистические термы $B_{i,j}$ ($i=\overline{1,\ m-1}$, $j=\overline{1, p_i}$, p_i – количество подинтервалов i-го вспомогательного фактора), описывающие значения вспомогательного фактора и представленные нечеткими множествами в виде:

Шаг 3. Выполним фаззификацию исторических данных. Определим подинтервал u_l ($l=\overline{1,\ n}$), которому принадлежит значение главного фактора.

1. Если значение главного фактора принадлежит интервалу u_1, то соответствующее ему фаззифицированное значение имеет вид:

$X_1 = 1/A_1 + 0{,}5/A_2$.

2. Если значение главного фактора принадлежит интервалу u_l, то соответствующее ему фаззифицированное значение имеет вид:

$X_l = 0{,}5/A_{i-1} + 1/A_i + 0{,}5/A_{i+1}$, $l=\overline{2,\ p-1}$.

3. Если значение главного фактора принадлежит интервалу u_n, то соответствующее ему фаззифицированное значение имеет вид:

$X_n = 0{,}5/A_{n-1} + 1/A_n$.

Определим подинтервал $v_{i,j}$ ($i=\overline{1,\ m-1}$, $j=\overline{1, p_\text{i}}$, p_i – количество подинтервалов i-го вспомогательного фактора), которому принадлежит значение i-го вспомогательного фактора по аналогии с u_l ($l=\overline{1,\ n}$).

Шаг 4. Построим m-факторную k-порядковую модель прогноза на основе нечетких временных серий, используя фаззифицированные главный и вспомогательный факторы (и соответствующие им фаззифицированные исторические данные, полученные на шаге 3).

Распределим нечеткие логические отношения в группы нечетких логических отношений на основе текущих состояний нечетких логических отношений. Вспомогательный фактор работает как вторая компонента в m-мерном векторе и будет использоваться на шаге 5.

Шаг 5. Вычислим прогнозируемые значения, используя следующие правила.

Искомое значение прогнозируемой величины F_{i+1} находится как сумма реального значения временного ряда (фактора) T_i для i-го периода и дефаззифицированного (четкого) значения приращения фактора y_{i+1}:

$F_{i+1} = T_i + y_{i+1}$.

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

$$y_{i+1} = \sum_{r=1}^{n} c_r \cdot w_r \Big/ \sum_{r=1}^{n} c_r\ , \tag{1}$$

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 соответствующую группу нечетких зависимостей.

Очевидно, что учет повторяющихся элементов (нечетких множеств) в правых частях нечетких логических групп позволил бы повысить точность прогноза.

Если при формировании групп нечетких зависимостей были выявлены повторяющиеся элементы в правых частях групп, то значение приращения y для ($i+1$)-го периода предлагается вычислять по формуле

$$y_{i+1} = \sum_{r=1}^{n} v_r \cdot w_r \Bigg/ \sum_{r=1}^{n} v_r ,$$

где w_r – дефаззифицированное значение (по формуле (1)), соответствующее нечеткому множеству X_r , v_r – количество повторений нечеткого множества x_r в правой части нечеткой логической зависимости, описывающей прогноз для ($i+1$)-го периода.

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

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

где F_i – предсказанное значение для i-го периода, T_i – реальное значение для i-го периода, m – количество значений (периодов) временного ряда.

Рассмотренная выше модель прогнозирования использует нечеткие множества первого типа для представления значений нечетких временных рядов. Для дальнейшего увеличения точности прогнозирования необходимо прибегнуть к использованию нечетких множеств второго типа. Нечеткие множества второго типа как расширение нечетких множеств первого типа были введены Л. Заде. Значение степени принадлежности для функции принадлежности нечеткого множества первого типа представляется обычным (четким) числом, в то время как значение степени принадлежности для функции принадлежности нечеткого множества второго типа представляется нечетким множеством. Преимущество нечетких множеств второго типа состоит в их способности выражать больше информации. Поэтому, заимствуя концепцию из нечетких множеств второго типа, можно разработать алгоритм, использующий большее количество информации при прогнозировании. В то же время алгоритмы, использующие нечеткие множества второго типа, отличаются большой вычислительной сложностью. Следовательно, необходимо использовать более простые вычислительные операции для уменьшения вычислительной сложности.

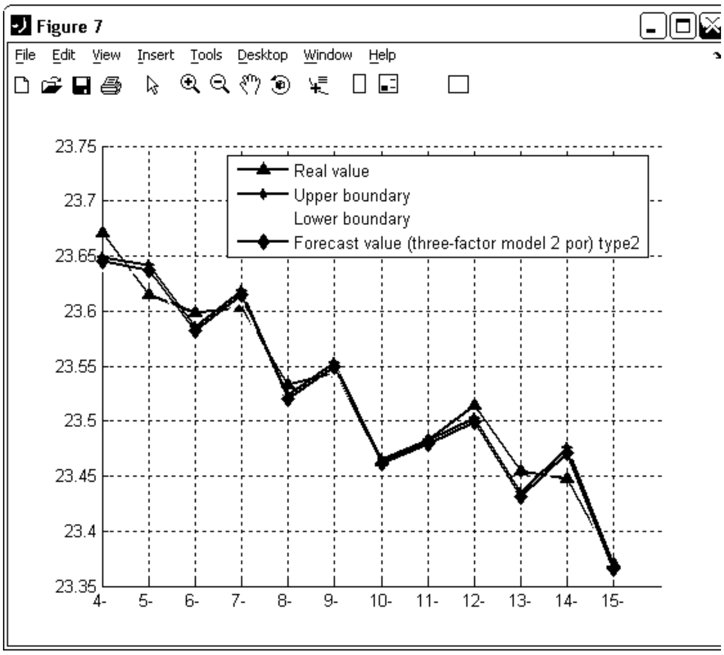
В частности, необходимо использовать операцию понижения типа для нечетких множеств второго типа. Понижение типа может быть достаточно трудоемкой операцией с вычислительной точки зрения, так как обычно нечеткое множество второго типа содержит большое число вложенных множеств первого типа. Однако для интервального нечеткого множества второго типа возможно применить широко используемый итерационный алгоритм Карника-Менделя, обеспечивающий поиск минимального и максимального центроидов вложенных нечетких множеств первого типа, которые в дальнейшем могут быть использованы при выполнении операции понижения типа (центроид нечетких множеств второго типа равен их среднему арифметическому) [4].

Кроме того, используя принцип расширения, можно использовать геометрическую интерпретацию центроида. Одна из причин использования такой интерпретации состоит в том, что невозможно выполнить понижение типа для геометрического множества, если внутри него содержится бесконечное число вложенных множеств. Используя геометрию, возможно вычислить центр площади геометрического нечеткого множества первого типа, эквивалентный центру площади дефаззификации. Эту же операцию можно расширить и на интервальные нечеткие множества второго типа.

Самостоятельной задачей при прогнозировании на основе нечетких временных рядов является определение оптимальных значений параметров модели, обеспечивающих максимальную точность прогнозирования.

В качестве примера было осуществлено прогнозирование курса валют. Во всех рассматриваемых случаях в качестве первого фактора использовалась цена одного доллара США по курсу ЦБ РФ, в качестве второго фактора использовалась лучшая цена покупки (за день) на торгах, в качестве третьего фактора – лучшая цена продажи (за день) на торгах.

Лучший результат был достигнут при прогнозировании с использованием трехфакторной модели второго порядка на основе нечетких множеств второго типа (рисунок).



Графическое отображение исходных и прогнозных данных для трехфакторной модели второго порядка на основе нечетких множеств второго типа

В этом случае средняя относительная ошибка прогнозирования составляет 0,043%. Из этого следует, что нечеткие множества второго типа обеспечивают увеличение точности прогноза, хотя при этом увеличивается вычислительная сложность алгоритма и время его реализации.

Полученные средние относительные ошибки прогнозирования свидетельствуют о пригодности разработанной модели к проведению краткосрочных прогнозов. Кроме того, как показал анализ, чем больше факторов будет учтено при прогнозировании, тем более точный прогноз можно получить.

Библиографический список

1. Демидова, Л.А. Применение нечетких временных серий для решения задачи прогнозирования эффективности инвестиционных вложений // Информационно-телекоммуникационные технологии: материалы 33-й межвузовской науч.-практ. конференции. – Рязань: Рязанское высшее военное командное училище связи, 2007. – С. 61 – 64.

2. Демидова, Л.А., Кираковский, В.В., Пылькин, А.Н. Алгоритмы и системы нечёткого вывода при решении задач диагностики городских инженерных коммуникаций в среде MATLAB. – М.: Радио и связь; Горячая линия – Телеком, 2005. – 365 с.

3. Chen, S.M. Forecasting enrollments based on fuzzy time series // Fuzzy Sets Systems. – 1996. – Vol. 81. –№ 3. – P. 311 – 319.

4. Mendel, J. M. Type-2 fuzzy sets and systems: an overview // IEEE Computational intellegence magazine. – 2007. – Vol. 2. – № 1. – P. 20 – 29.