ТЕОРЕТИЧНІ ТА ПРИКЛАДНІ ПРОБЛЕМИ ІНТЕЛЕКТУАЛЬНИХ СИСТЕМ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ

УДК 618.324

НЕЧЕТКИЙ МЕТОД ГРУППОВОГО УЧЕТА АРГУМЕНТОВ ПРИ НЕОПРЕДЕЛЕННЫХ ВХОДНЫХ ДАННЫХ

Ю.П. ЗАЙЧЕНКО

Предложено развитие нечеткого метода группового учета аргументов (НМГУА) для случая, когда входные переменные заданы нечетко в виде интервалов неопределенности. Построена математическая модель задачи нахождения оптимальной модели при нечетких исходных данных. Предложен алгоритм НМГУА с нечеткими входами. Приведены результаты экспериментальных исследований этого алгоритма в задаче прогнозирования на фондовом рынке.

ВСТУПЛЕНИЕ

Нечеткий метод группового учета аргументов (НМГУА) с нечеткими входными данными является модификацией нечеткого метода группового учета аргументов и использует его основные идеи. Суть метода заключается в построении неизвестной функциональной зависимости между входными и выходными данными, когда входные переменные заданы нечетко в виде интервалов неопределенности. Для этого на каждом этапе строятся модели на основе скрещивания пар входных переменных, выбирается определенное количество наилучших и передается на следующий этап (ряд селекции). Процесс завершается генерацией оптимальной модели [1].

Модифицированный НМГУА, который оперирует с нечеткими входными данными, использует алгоритм нечеткого метода группового учета аргументов и другие математические модели.

ОБЩИЙ ВИД МАТЕМАТИЧЕСКОЙ МОДЕЛИ НМГУА С НЕЧЕТКИМИ ВХОДНЫМИ ДАННЫМИ

Рассмотрим линейную интервальную модель регрессии

$$Y = A_0 Z_0 + A_1 Z_1 + \ldots + A_n Z_n, \tag{1}$$

где A_i — нечеткие числа, которые описываются тремя параметрами $A_i = (\underline{A_i}, \breve{A}_i, \overline{A_i})$, где $\breve{A}_i$ — центр интервала, $\overline{A_i}$ — его верхняя граница,

$\underline{A_i}$ — нижняя граница; Z_i — также нечеткие числа, задаваемые параметрами $(\underline{Z_i}, \breve{Z}_i, \overline{Z_i})$, где $\underline{Z_i}$ — нижняя граница, $\breve{Z}_i$ — центр, $\overline{Z_i}$ — верхняя граница нечеткого числа.

Тогда Y — нечеткое число, параметры которого определяются следующим образом (в соответствии с формулами для умножения L - R -чисел):

центр интервала

$$\breve{y} = \sum \breve{A}_i * \breve{Z}_i ;$$

отклонение в левой части функции принадлежности

$$\breve{y} - \underline{y} = \sum_i \left(\left| \breve{A}_i \right| * (\breve{Z}_i - \underline{Z}_i) + (\breve{A}_i - \underline{A_i}) * \left| \breve{Z}_i \right| \right),$$

откуда нижняя граница интервала

$$\underline{y} = \sum (\breve{A}_i * \breve{Z}_i - \left| \breve{A}_i \right| * (\breve{Z}_i - \underline{Z}_i) - (\breve{A}_i - \underline{A_i}) * \left| \breve{Z}_i \right|) ; \qquad (2)$$

отклонение в правой части функции принадлежности

$$\overline{y} - \breve{y} = \sum \left(\left| \breve{A}_i \right| * (\overline{Z}_i - \breve{Z}_i) + \left| \breve{Z}_i \right| * (\overline{A}_i - \breve{A}_i) \right),$$

откуда верхняя граница интервала

$$\overline{y} = \sum \left(\left| \breve{A}_i \right| * (\overline{Z}_i - \breve{Z}_i) + \left| \breve{Z}_i \right| * (\overline{A}_i - \breve{A}_i) + \breve{A}_i * \breve{Z}_i \right). \qquad (3)$$

Требуется построить оптимальную модель, позволяющую работать с нечеткими входными переменными Z_i приведенного выше вида. Для корректности интервальной модели необходимо, чтобы действительное значение исходной величины Y принадлежало полученному в результате работы метода интервалу.

Следовательно, основные требования к оценочной линейной интервальной модели заключаются в нахождении таких значений параметров $(\underline{A_i}, \breve{A}_i, \overline{A_i})$ нечетких коэффициентов, при которых:

а) наблюдаемые значения y_k попадали бы в оценочный интервал для Y_k;

б) суммарная ширина оценочного интервала была бы минимальной.

Входными данными в этой задаче являются: $Z_k = [Z_{ki}]_i$ — входная обучающая выборка, а также y_k — известные нам исходные значения, $k = \overline{1, M}$, M — количество точек наблюдения.

В данной работе рассматриваются два случая вида функции принадлежности (ФП) нечетких величин:

1) треугольные ФП;
2) ФП Гаусса.

Вид частичных описаний выбран квадратичным.

$$f(x_i, x_j) = A_0 + A_1 x_i + A_2 x_j + A_3 x_i x_j + A_4 x_i^2 + A_5 x_j^2 .$$

МОДЕЛЬ НМГУА С НЕЧЕТКИМИ ВХОДНЫМИ ДАННЫМИ ДЛЯ ТРЕУГОЛЬНЫХ ФП

Вид математической модели для случая треугольных ФП

Рассмотрим линейную интервальную модель регрессии

$$Y = A_0 Z_0 + A_1 Z_1 + \ldots + A_n Z_n ,$$

где A_i — нечеткие числа треугольной формы, которые описываются тройкой параметров $A_i = (\underline{A_i}, a_i, \overline{A_i})$, где a_i — центр интервала; $\overline{A_i}$ — его верхняя граница; $\underline{A_i}$ — нижняя граница.

В данной задаче рассматривается случай симметричных ФП для параметров A_i, и их можно описать парой параметров (a_i , c_i) (рис. 1).

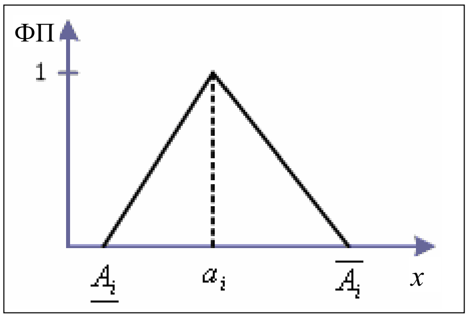

Рис. 1. Треугольная ФП

$\underline{A}_i = a_i - c_i$; $\overline{A}_i = a_i + c_i$, c_i — ширина интервала, $c_i \geq 0$; Z_i — также нечеткие числа треугольной формы, которые задаются параметрами $(\underline{Z_i}, \breve{Z}_i, \overline{Z_i})$, $\underline{Z_i}$ — нижняя граница, $\breve{Z}_i$ — центр, $\overline{Z_i}$ — верхняя граница нечеткого числа.

Тогда Y — нечеткое число, параметры которого определятся следующим образом:

центр интервала

$$\breve{y} = \sum_i a_i * \breve{Z}_i ;$$

отклонение в левой части ФП

$$\breve{y} - \underline{y} = \sum_i |a_i| * (\breve{Z}_i - \underline{Z}_i) + c_i |\breve{Z}_i| ,$$

откуда нижняя граница интервала

$$\underline{y} = \sum_i a_i * \breve{Z}_i - |a_i| * (\breve{Z}_i - \underline{Z}_i) - c_i |\breve{Z}_i| ;$$

отклонение в правой части ФП

$$\overline{y} - \breve{y} = \sum_i |a_i| * (\overline{Z}_i - \breve{Z}_i) + c_i |\breve{Z}_i| ,$$

откуда верхняя границя интервала

$$\overline{y} = \sum_i |a_i| * (\overline{Z}_i - \breve{Z}_i) + c_i |\breve{Z}_i| + a_i * \breve{Z}_i .$$

Для корректности интервальной модели необходимо, чтобы действительное значение выходной величины Y принадлежало полученному в результате работы метода интервалу. Это можно записать так:

$$\begin{cases} \sum_i a_i * \breve{Z}_i - |a_i| \ (\breve{Z}_i - \underline{Z}_i) - c_i \left|\breve{Z}_i\right| \le y_k , \\ \sum_i a_i * \breve{Z}_i + |a_i| \ (\overline{Z}_i - \breve{Z}_i) + c_i \left|\breve{Z}_i\right| \ge y_k , \quad k = \overline{1,M}, \end{cases}$$

где $Z_k = [Z_k]_i$ — входная обучающая выборка; y_k — известные нам исходные значения; $k = \overline{1,M}$, M — количество точек наблюдения.

Следовательно, основные требования к оценочной линейной интервальной модели для треугольного частичного описания заключаются в нахождении таких значений параметров (a_i, c_i) нечетких коэффициентов, при которых:

а) наблюдаемые значения y_k попадали бы в оценочный интервал для Y_k;

б) суммарная ширина оценочного интервала была бы минимальной.

Эти требования можно свести к задаче линейного программирования

$$\min_{a_i, c_i} \sum_{k=1}^{M} |a_i| \cdot \left(\overline{Z}_i - \underline{Z}_i\right) + 2c_i \left|\breve{Z}_i\right| \tag{4}$$

при условиях

$$\begin{cases} \sum_i a_i * \breve{Z}_i - |a_i| \cdot (\breve{Z}_i - \underline{Z}_i) - c_i \left|\breve{Z}_i\right| \le y_k , \\ \sum_i a_i * \breve{Z}_i + |a_i| \cdot (\overline{Z}_i - \breve{Z}_i) + c_i \left|\breve{Z}_i\right| \ge y_k , \quad k = \overline{1,M}. \end{cases} \tag{5}$$

Формализованная постановка задачи для случая треугольных ФП

Рассмотрим частичное описание вида

$$f(x_i, x_j) = A_0 + A_1 x_i + A_2 x_j + A_3 x_i x_j + A_4 x_i^2 + A_5 x_j^2 \ . \tag{6}$$

Запишем его в соответствии с моделью (1). Для этого в ней нужно предположить, что $z_0 = 1$, $z_1 = x_i$, $z_2 = x_j$, $z_3 = x_i x_j$, $z_4 = x_i^2$, $z_5 = x_j^2$.

Тогда математическая модель (4), (5) запишется в следующем виде:

$$\min_{a_i, c_i} \Bigg((2Mc_0 + |a_1| \sum_{k=1}^{M} (\overline{x}_{ik} - \underline{x}_{ik}) + 2c_1 \sum_{k=1}^{M} \left|\breve{x}_{ik}\right| + |a_2| \sum_{k=1}^{M} (\overline{x}_{jk} - \underline{x}_{jk}) + 2c_2 \sum_{k=1}^{M} \left|\breve{x}_{jk}\right| +$$

$$+ |a_3| \sum_{k=1}^{M} (\left|\breve{x}_{ik}\right| (\overline{x}_{jk} - \underline{x}_{jk}) + \left|\breve{x}_{jk}\right| (\overline{x}_{ik} - \underline{x}_{ik})) + 2c_3 \sum_{k=1}^{M} \left|\breve{x}_{ik} \breve{x}_{jk}\right| + \tag{7}$$

$$+ 2|a_4| \sum_{k=1}^{M} \left|\breve{x}_{ik}\right| (\overline{x}_{ik} - \underline{x}_{ik}) + 2c_4 \sum_{k=1}^{M} \breve{x}_{ik}^2 + 2|a_5| \sum_{k=1}^{M} \left|\breve{x}_{jk}\right| (\overline{x}_{jk} - \underline{x}_{jk}) + 2c_5 \sum_{k=1}^{M} \breve{x}_{jk}^2 \Bigg)$$

при условиях

$$a_0 + a_1 \underline{x}_{ik} + a_2 \underline{x}_{jk} + a_3 (-\left|\breve{x}_{ik}\right| (\breve{x}_{jk} - \underline{x}_{jk}) - \left|\breve{x}_{jk}\right| (\breve{x}_{ik} - \underline{x}_{ik}) + \breve{x}_{ik} \breve{x}_{jk}) +$$

$$+ a_4 (-2\left|\breve{x}_{ik}\right| (\breve{x}_{ik} - \underline{x}_{ik}) + \breve{x}_{ik}{}^2) + a_5 (2\left|\breve{x}_{jk}\right| (\breve{x}_{jk} - \underline{x}_{jk}) + \breve{x}_{jk}{}^2) - c_0 - c_1 \left|\breve{x}_{ik}\right| - \tag{8}$$

$$-c_2\left|\breve{x}_{jk}\right| - c_3\left|\breve{x}_{ik}\breve{x}_{jk}\right| - c_4\breve{x}_{ik}^2 - c_5\breve{x}_{jk}^2 \le y_k\,,$$

$$a_0 + a_1\breve{x}_{ik} + a_2\breve{x}_{jk} + a_3(\left|\breve{x}_{ik}\right|(\bar{x}_{jk} - \breve{x}_{jk}) + \left|\breve{x}_{jk}\right|(\bar{x}_{ik} - \breve{x}_{ik}) - \breve{x}_{ik}\breve{x}_{jk}) +$$

$$+\, a_4(2\left|\breve{x}_{ik}\right|(\bar{x}_{ik} - \breve{x}_{ik}) - \breve{x}_{ik}{}^2) + a_5(2\left|\breve{x}_{jk}\right|(\bar{x}_{jk} - \breve{x}_{jk}) - \breve{x}_{jk}{}^2) +$$

$$+\, c_0 + c_1\left|\breve{x}_{ik}\right| + c_2\left|\breve{x}_{jk}\right| + c_3\left|\breve{x}_{ik}\breve{x}_{jk}\right| + c_4\breve{x}_{ik}^2 + c_5\breve{x}_{jk}^2 \ge y_k\,,$$

$$c_i \ge 0,\ \ i = \overline{0{,}5}\,.$$

Как видно, эта задача является задачей линейного программирования, но поскольку нет ограничений неотрицательности для переменных a_i, то для ее решения переходим к двойственной задаче, вводя двойственные переменные $\{\delta_k\}$ и $\{\delta_{k+M}\}$.

Запишем двойственную задачу

$$\max\left(\sum_{k=1}^{M} y_k\,\delta_{k+M} - \sum_{k=1}^{M} y_k\,\delta_k\right) \tag{9}$$

при условиях

$$\sum_{k=1}^{M}\delta_{k+M} - \sum_{k=1}^{M}\delta_k = 0\,,$$

$$\sum_{k=1}^{M}\bar{x}_{ik}\,\delta_{k+M} - \sum_{k=1}^{M}\underline{x}_{ik}\,\delta_k = \sum_{k=1}^{M}(\bar{x}_{ik} - \underline{x}_{ik})\,,$$

$$\sum_{k=1}^{M}\bar{x}_{jk}\,\delta_{k+M} - \sum_{k=1}^{M}\underline{x}_{jk}\,\delta_k = \sum_{k=1}^{M}(\bar{x}_{jk} - \underline{x}_{jk})\,,$$

$$\sum_{k=1}^{M}(\left|\breve{x}_{ik}\right|(\bar{x}_{jk} - \breve{x}_{jk}) + \left|\breve{x}_{jk}\right|(\bar{x}_{ik} - \breve{x}_{ik}) - \breve{x}_{ik}\breve{x}_{jk})\delta_{k+M} -$$

$$-\sum_{k=1}^{M}(-\left|\breve{x}_{ik}\right|(\breve{x}_{jk} - \underline{x}_{jk}) - \left|\breve{x}_{jk}\right|(\breve{x}_{ik} - \underline{x}_{ik}) + \breve{x}_{ik}\breve{x}_{jk}) * \delta_k =$$

$$= \sum_{k=1}^{M}(\left|\breve{x}_{ik}\right|(\bar{x}_{jk} - \underline{x}_{jk}) + \left|\breve{x}_{jk}\right|(\bar{x}_{ik} - \underline{x}_{ik}))\,, \tag{10}$$

$$\sum_{k=1}^{M}(2\left|\breve{x}_{ik}\right|(\bar{x}_{ik} - \breve{x}_{ik}) - \breve{x}_{ik}{}^2)\delta_{k+M} - \sum_{k=1}^{M}(-2\left|\breve{x}_{ik}\right|(\breve{x}_{ik} - \underline{x}_{ik}) + \breve{x}_{ik}{}^2)\delta_k =$$

$$= \sum_{k=1}^{M}\left|\breve{x}_{jk}\right|(\bar{x}_{jk} - \underline{x}_{jk})\,,$$

$$\sum_{k=1}^{M}\delta_{k+M} + \sum_{k=1}^{M}\delta_k \le 2M\,,$$

$$\sum_{k=1}^{M} \left|\breve{x}_{ik}\right| \delta_{k+M} + \sum_{k=1}^{M} \left|\breve{x}_{ik}\right| \delta_k \le 2 \sum_{k=1}^{M} \left|\breve{x}_{ik}\right|,$$

$$\sum_{k=1}^{M} \left|\breve{x}_{jk}\right| \delta_{k+M} + \sum_{k=1}^{M} \left|\breve{x}_{jk}\right| \delta_k \le 2 \sum_{k=1}^{M} \left|\breve{x}_{jk}\right|,$$

$$\sum_{k=1}^{M} \left|\breve{x}_{ik} \breve{x}_{jk}\right| \delta_{k+M} + \sum_{k=1}^{M} \left|\breve{x}_{ik} \breve{x}_{jk}\right| \delta_k \le 2 \sum_{k=1}^{M} \left|\breve{x}_{ik} \breve{x}_{jk}\right|, \tag{11}$$

$$\sum_{k=1}^{M} \breve{x}_{ik}^2 \delta_{k+M} + \sum_{k=1}^{M} \breve{x}_{ik}^2 \delta_k \le 2 \sum_{k=1}^{M} \breve{x}_{ik}^2,$$

$$\sum_{k=1}^{M} \breve{x}_{jk}^2 \delta_{k+M} + \sum_{k=1}^{M} \breve{x}_{jk}^2 \delta_k \le 2 \sum_{k=1}^{M} \breve{x}_{jk}^2,$$

$$\delta_k \ge 0, \quad \delta_{k+M} \ge 0, \quad k = \overline{1, M}.$$

Решив двойственную задачу (9)–(11) симплекс-методом и найдя оптимальные значения двойственных переменных $\{\delta_k\}, \{\delta_{k+M}\}$, найдем оптимальные значения искомых переменных c_i, a_i, $i = \overline{0,5}$, а также искомую нечеткую модель для заданного частичного описания.

НМГУА с нечеткими входными данными для ФП Гаусса

Рассмотрим теперь реализацию НМГУА с нечеткими входами для случая гауссовских ФП (рис. 2).

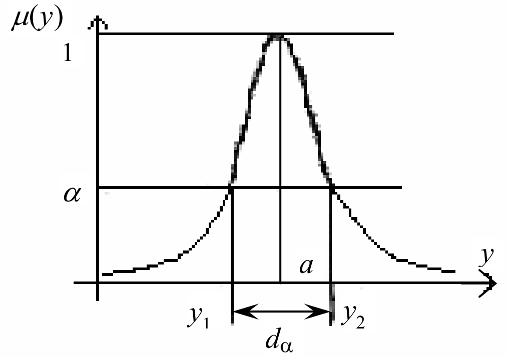

Рис. 2. ФП Гаусса. Множество уровня *α*

Вид математической модели для случая ФП Гаусса

Рассмотрим линейную интервальную модель регрессии

$$Y = A_0 Z_0 + A_1 Z_1 + \ldots + A_n Z_n,$$

где A_i — нечеткие числа гауссовской формы, которые описываются парой параметров (a_i, c_i), где a_i — центр; c_i — величина, характеризующая ширину интервала, $c_i \ge 0$; Z_i — нечеткие числа гауссовской формы, имеющие ФП следующего вида:

$$\mu(z) = \begin{cases} e^{-\frac{1}{2}\frac{(z-a)^2}{c_1^2}}, & z \le a, \\ e^{-\frac{1}{2}\frac{(z-a)^2}{c_2^2}}, & z > a. \end{cases} \qquad (12)$$

Тогда такие нечеткие числа описываются тройками параметров (c_1, a, c_2), где a — центр; c_1 — величина, характеризующая ширину интервала в левой ветви ФП; c_2 — в правой.

В этом случае задача формулируется следующим образом: найти такие нечеткие параметры (a_i, c_i) нечетких коэффициентов A_i, чтобы выполнялись условия [2]:

1) наблюдение y_k принадлежит к оценочному интервалу Y_k со степенью, не меньшей , чем α, $0 < \alpha < 1$, $k = \overline{1, M}$;

2) ширина оценочного интервала уровня α минимальна.

Ширина оценочного интервала уровня α равняется

$$d_\alpha = y_2 - y_1. \qquad (13)$$

Найти y_1 и y_2 можно из системы

$$\begin{cases} \alpha = \exp\left\{-\frac{1}{2}\frac{(y_2 - a)^2}{c_2^2}\right\}, \\ \alpha = \exp\left\{-\frac{1}{2}\frac{(y_1 - a)^2}{c_1^2}\right\}. \end{cases} \qquad (14)$$

Откуда

$$\begin{cases} y_2 = c_2\sqrt{-2\ln\alpha} + a, \\ y_1 = -c_1\sqrt{-2\ln\alpha} + a, \end{cases}$$

а $d_\alpha = \sqrt{-2\ln\alpha}(c_2 + c_1)$.

Условие 1) можно записать таким образом: $\mu(y_k) \ge \alpha$, что можно свести к системе

$$\begin{cases} y_k \le a_k + c_{2k}\sqrt{-2\ln\alpha}, \\ y_k \ge a_k - c_{1k}\sqrt{-2\ln\alpha}. \end{cases}$$

Следовательно, задачу можно записать в виде

$$\min\sum_{k=1}^{M} d_\alpha^k = \min\sum_{k=1}^{M}(c_{2k} + c_{1k})\sqrt{-2\ln\alpha}, \qquad (15)$$

$$\begin{cases} y_k \le a_k + c_{2k}\sqrt{-2\ln\alpha}, \\ y_k \ge a_k - c_{1k}\sqrt{-2\ln\alpha}. \end{cases} \qquad (16)$$

Переменные Z_i можно также задавать параметрами $(\underline{Z_i}, \breve{Z}_i, \overline{Z_i})$, $\underline{Z_i}$ — нижняя граница; $\breve{Z}_i$ — центр; $\overline{Z_i}$ — верхняя граница нечеткого числа, где

$$\underline{Z_i} = \breve{Z}_i - c_{1i}, \quad \overline{Z_i} = \breve{Z}_i + c_{2i}.$$

Тогда Y — нечеткое число, параметры которого определяются соотношениями, аналогичными (2), (3).

Формализованная постановка задачи для случая ФП Гаусса

Рассмотрим частичное описание вида

$$f(x_i, x_j) = A_0 + A_1 x_i + A_2 x_j + A_3 x_i x_j + A_4 x_i^2 + A_5 x_j^2.$$

Запишем его в соответствии с моделью (1). Для этого в ней нужно положить, что

$$z_0 = 1,\ z_1 = x_i,\ z_2 = x_j,\ z_3 = x_i x_j,\ z_4 = x_i^2,\ z_5 = x_j^2.$$

Тогда математическая модель (7), (8) запишется в следующем виде:

$$\min_{a_i, c_i} \Big(2Mc_0 + a_1 \sum_{k=1}^{M} (\overline{x}_{ik} - \underline{x}_{ik}) + 2c_1 \sum_{k=1}^{M} |\breve{x}_{ik}| + a_2 \sum_{k=1}^{M} (\overline{x}_{jk} - \underline{x}_{jk}) + 2c_2 \sum_{k=1}^{M} |\breve{x}_{jk}| +$$

$$+ a_3 \sum_{k=1}^{M} (|\breve{x}_{ik}| (\overline{x}_{jk} - \underline{x}_{jk}) + |\breve{x}_{jk}| (\overline{x}_{ik} - \underline{x}_{ik})) + 2c_3 \sum_{k=1}^{M} |\breve{x}_{ik} \breve{x}_{jk}| + \tag{17}$$

$$+ 2a_4 \sum_{k=1}^{M} |\breve{x}_{ik}| (\overline{x}_{ik} - \underline{x}_{ik}) + 2c_4 \sum_{k=1}^{M} \breve{x}_{ik}^2 + 2a_5 \sum_{k=1}^{M} |\breve{x}_{jk}| (\overline{x}_{jk} - \underline{x}_{jk}) + 2c_5 \sum_{k=1}^{M} \breve{x}_{jk}^2 \Big)$$

при условиях

$$a_0 + a_1(\breve{x}_{ik} - \sqrt{-2\ln\alpha}(\breve{x}_{ik} - \underline{x}_{ik})) + a_2(\breve{x}_{jk} - \sqrt{-2\ln\alpha}(\breve{x}_{jk} - \underline{x}_{jk})) + a_3(\breve{x}_{ik}\breve{x}_{jk} -$$

$$- (|\breve{x}_{ik}|(\breve{x}_{jk} - \underline{x}_{jk}) + |\breve{x}_{jk}|(\breve{x}_{ik} - \underline{x}_{ik}))\sqrt{-2\ln\alpha}) +$$

$$+ a_4(\breve{x}_{ik}{}^2 - 2\sqrt{-2\ln\alpha}(|\breve{x}_{ik}|(\breve{x}_{ik} - \underline{x}_{ik})) +$$

$$+ a_5(\breve{x}_{jk}{}^2 - 2\sqrt{-2\ln\alpha}(|\breve{x}_{jk}|(\breve{x}_{jk} - \underline{x}_{jk}))) -$$

$$- c_0\sqrt{-2\ln\alpha} - c_1|\breve{x}_{ik}|\sqrt{-2\ln\alpha} - c_2|\breve{x}_{jk}|\sqrt{-2\ln\alpha} -$$

$$- c_3|\breve{x}_{ik}\breve{x}_{jk}|\sqrt{-2\ln\alpha} - c_4\breve{x}_{ik}^2\sqrt{-2\ln\alpha} - c_5\breve{x}_{jk}^2\sqrt{-2\ln\alpha} \le y_k, \tag{18}$$

$$a_0 + a_1(\breve{x}_{ik} + \sqrt{-2\ln\alpha}(\overline{x}_{ik} - \breve{x}_{ik})) + a_2(\breve{x}_{jk} + \sqrt{-2\ln\alpha}(\overline{x}_{jk} - \breve{x}_{jk})) +$$

$$+a_3(\breve{x}_{ik}\breve{x}_{jk}+(|\breve{x}_{ik}|(\bar{x}_{jk}-\breve{x}_{jk})+|\breve{x}_{jk}|(\bar{x}_{ik}-\breve{x}_{ik}))\sqrt{-2\ln\alpha})+$$

$$+a_4(\breve{x}_{ik}{}^2+2\sqrt{-2\ln\alpha}(|\breve{x}_{ik}|(\bar{x}_{ik}-\breve{x}_{ik}))+$$

$$+a_5(\breve{x}_{jk}{}^2+2\sqrt{-2\ln\alpha}(|\breve{x}_{jk}|(\bar{x}_{jk}-\breve{x}_{jk}))+c_0\sqrt{-2\ln\alpha}+$$

$$+c_1|\breve{x}_{ik}|\sqrt{-2\ln\alpha}+c_2|\breve{x}_{jk}|\sqrt{-2\ln\alpha}+c_3|\breve{x}_{ik}\breve{x}_{jk}|\sqrt{-2\ln\alpha}+$$

$$+c_4\breve{x}_{ik}^2\sqrt{-2\ln\alpha}+c_5\breve{x}_{jk}^2\sqrt{-2\ln\alpha}\geq y_k\,,\ \ c_i\geq 0,\ \ i=\overline{0{,}5}\,. \tag{19}$$

Как видно, эта задача является задачей линейного программирования (ЛП), но поскольку нет ограничений неотрицательности для переменных a_i, то для ее решения переходим к двойственной задаче, вводя двойственные переменные $\{\delta_k\}$ и $\{\delta_{k+M}\}$.

Решив симплекс-методом задачу, двойственную к (17) – (19), и найдя оптимальные значения двойственных переменных $\{\delta_k\}$, $\{\delta_{k+M}\}$, определим оптимальные значения искомых переменных c_i, a_i, $i=\overline{0{,}5}$, а также искомую нечеткую модель для частичного описания.

Описание алгоритма НМГУА

1. Выбор общего вида модели, которая будет описывать искомую зависимость.

2. Выбор внешних критериев оптимальности: критерия регулярности $\bar{\delta}^2$ или несмещенности $N_{\text{см}}$, где

$$\bar{\delta}^2=\frac{1}{N_{\text{пров}}}\sum_{i=1}^{N_{\text{пров}}}\left(y_i-\overline{y_i}(x)\right)^2;$$

y_i — реальные данные; $\overline{y}_i(x)$ — выход модели (прогноз); $N_{\text{пров}}$ — объем проверочной выборки.

$$N_{\text{см}}=\frac{1}{N_1+N_2}\sum_{i=1}^{N}\left(y_i^*-y_i^{**}\right)^2,$$

где y_i^* — выход модели 1, построенной на выборке N_1; y_i^{**} — выход модели 2, построенной на выборке N_2.

3. Выбор общего вида опорной функции (частичных описаний), например, линейного или квадратичного.

4. Разбиение выборки на обучающую $N_{\text{об}}$ и проверочную $N_{\text{пров}}$.

5. Присвоение нулевых значений счетчику числа моделей k и счетчику числа рядов r.

6. Генерация новой частичной модели f_r вида (6) на обучающей выборке. Решение задачи ЛП (9), (10) и нахождение искомых значений a_i и c_i.

7. Определение по проверочной выборке $N_{\text{пров}}$ значения внешнего критерия $\left(N_{\text{см}k}^{(r)}\right)$ или $\left(\overline{\delta}_k^{(2)}(k)\right)$ для k-го частичного описания r-й итерации.

8. $k = k+1$. Если $k \geq C_F^{(2)}$, то $k = 0$, $r = r+1$.

9. Вычисление среднего критерия для моделей r-й итерации ($\left(N_{\text{см}}^{(r)}\right)$ или $\delta^{(2)}(r)$). Если $r = 1$, то переходим на шаг 6, иначе — на шаг 10.

10.Если $\left|N_{\text{см}}^{(r)} - N_{\text{см}}^{(r-1)}\right| \leq \varepsilon$, то переходим на шаг 11, иначе — отбираем F лучших моделей, затем, положив $r = r+1$, $k = 1$, переходим на шаг 6 и выполняем следующую $(r+1)$-ю итерацию.

11.Из F моделей предыдущего ряда находим по критерию регуляризации наилучшую модель.

ЭКСПЕРИМЕНТАЛЬНЫЕ ИССЛЕДОВАНИЯ НМГУА И АНАЛИЗ ПОЛУЧЕННЫХ РЕЗУЛЬТАТОВ

Прогнозирование индекса РТС. Общая информация

НП Фондовая биржа «Российская торговая система» (РТС) была создана в 1995 г. с целью объединить разрозненные региональные рынки в единый рынок ценных бумаг России.

Информация относительно торгов в РТС является важнейшим источником данных о состоянии российского рынка ценных бумаг, поскольку именно она обслуживает значительную часть иностранных и российских портфельных инвестиций в акциях российских компаний. РТС — общепризнанный центр ценообразования ценных бумаг широкого круга эмитентов.

К торгам в РТС допущено около 270 ценных бумаг, в том числе более 43 облигаций. В информационных системах содержится информация относительно индикативных котировок около 750 акций и 500 векселей российских компаний.

Индекс РТС — официальный индикатор Фондовой биржи РТС. Он рассчитывается на протяжении торговой сессии при каждом изменении цен акций, включенных в список для его расчета. Первое значение индекса — это значение открытия, последнее — закрытия.

Индекс рассчитывается в двух значениях — валютном и рублевом. Рублевые значения являются вспомогательными и рассчитываются на основе валютных значений.

Индекс рассчитывается при возникновении каких-либо событий с любыми ценными бумагами, которые входят в список для его расчета.

Входные данные для расчета индекса:

- Информация о соглашениях, которые были заключены в торговой системе на протяжении торговой сессии с ценными бумагами, которые входят в список для расчета индекса и имеют объем не меньший, чем объем, требуемый в правилах торговли к заявкам по данным ценным бумагам. При этом цена соглашения должна быть не ниже цены лучшей стандартной заявки на приобретение и не выше цены лучшей стандартной заявки на продажу.
- Информация о текущих лучших ценах стандартных заявок.

Значения индекса рассчитываются с точностью до двух знаков после запятой. Значения цен ценных бумаг — с точностью до пяти знаков после запятой.

Список ценных бумаг для расчета индекса РТС складывается из наиболее ликвидных акций российских компаний, отобранных Информационным комитетом на основе экспертных оценок. Количество ценных бумаг в индексе не должно превышать 50.

Раскрытие информации об индексе происходит на веб-сайте РТС на странице www.rts.ru/rtsindex.

Результаты работы НМГУА с нечеткими входными данными при прогнозировании индекса РТС

Эксперимент 1. Прогнозирование индекса РТС (цена открытия)

Использовалось пять нечетких входных переменных, которые представляют собой цены на акции ведущих энергетических компаний России, включенных в список для расчета индекса РТС:

LKOH — акции ОАО ЛУКОЙЛ, EESR — акции ОАО РАО ЕЭС России, YUKO — акции ОАО ЮКОС, SNGSP — привилегированные акции ОАО «Сургутнефтегаз», SNGS — обычные акции ОАО «Сургутнефтегаз».

Исходной переменной служили значения индекса РТС (цена открытия) за тот же период (03.04.2006 – 18.05.2006).

Размер выборки — 32 значения.

Размер обучающей выборки — 18 значений (оптимальный размер обучающей выборки для данного эксперимента).

Получены следующие результаты.

- Для треугольного вида ФП

1. Для нормированных входных данных (рис. 3). Нормирование выполнялось по формуле

$$X_i = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}}, \quad X_i \in [0;1].$$

Значение критерия составило: $\text{СКО} = 0{,}055557$.

2. Для ненормированных входных данных значения критериев составили: $\text{СКО} = 18{,}48657$, $\text{MAPE (средняя относительная ошибка)} = 0{,}8\%$ (рис. 4).

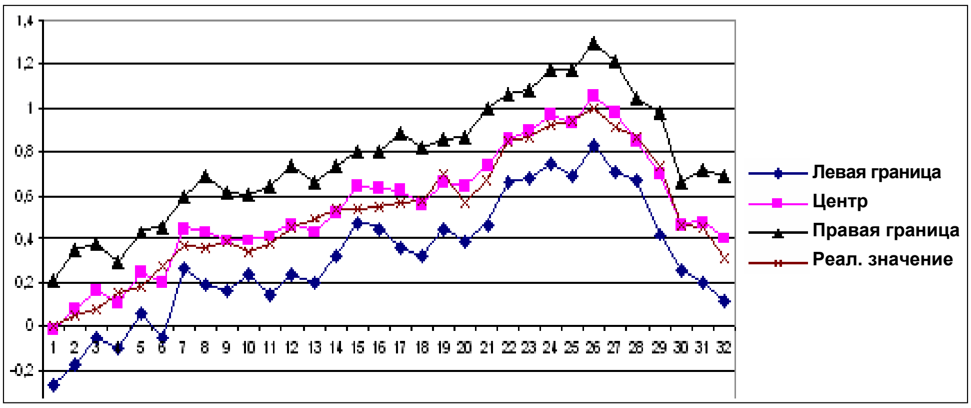

Рис. 3. Результаты эксперимента 1 для треугольной ФП и нормированных значений входных переменных

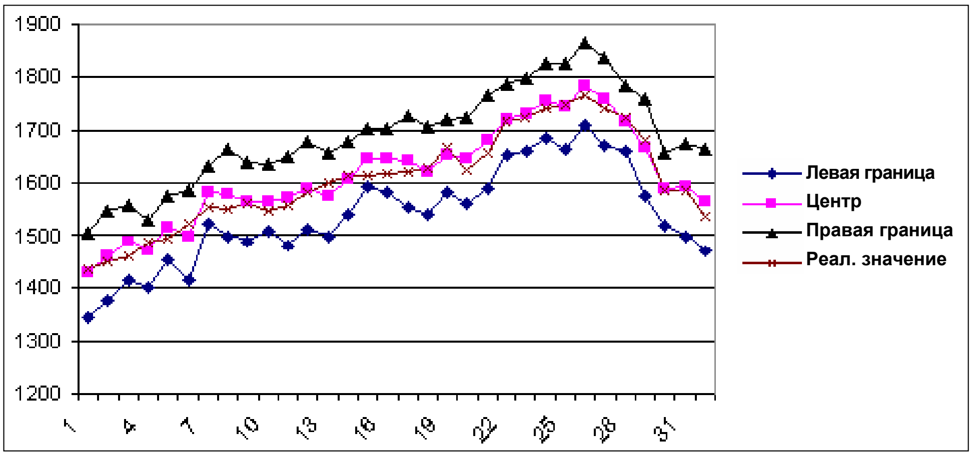

Рис. 4. Результаты эксперимента 1 для треугольной ФП и ненормированных значений входных переменных

- Для случая ФП Гаусса (оптимальный уровень $\alpha = 0{,}8$)

1. Для нормированных входных данных значение критерия составило: СКО $= 0{,}028013$ (рис. 5).

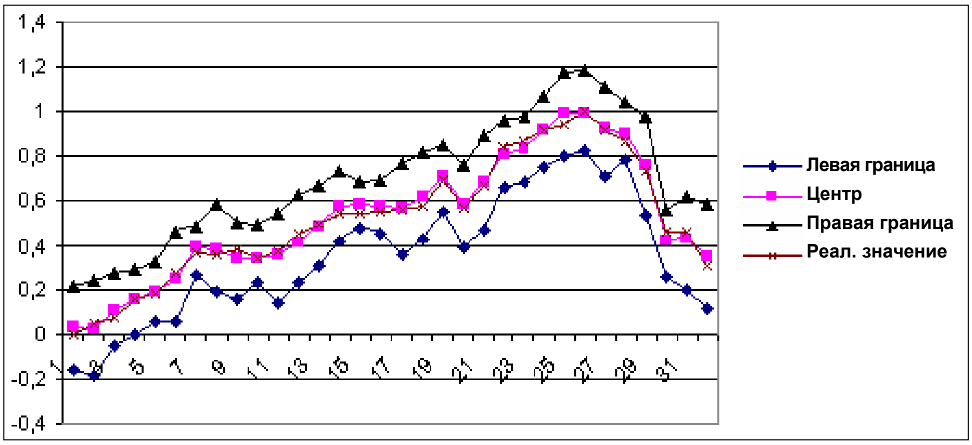

Рис. 5. Результаты эксперимента 1 для ФП Гаусса и нормированных значений входных переменных

2. Для ненормированных входных данных значения критериев составили: $СКО = 9{,}321461$, $MAPE = 0{,}4\%$ (рис. 6).

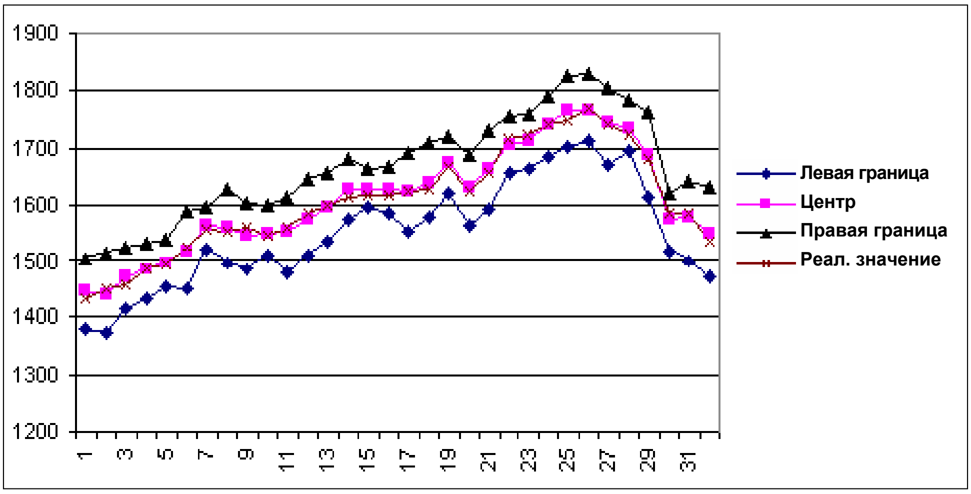

Рис. 6. Результаты эксперимента 1 для ФП Гаусса и ненормированных значений входных переменных

Как показывает эксперимент 1, прогнозирование с использованием треугольных и гауссовских ФП дает хорошие результаты, причем результаты экспериментов с ФП Гаусса лучше, чем с треугольной ФП.

Для нормированных данных

Критерии	Треугольные ФП	ФП Гаусса
СКО	0,055557	0,028013

Для ненормированных данных

Критерии	Треугольные ФП	ФП Гаусса
СКО	18,48657	9,321461
MAPE	0,8%	0,4%

Эксперимент 2. Прогнозирование индекса РТС (цена закрытия)

Использовались те же входные переменные, что и в эксперименте 1. Исходной переменной были значения индекса РТС (цена закрытия) за тот же период (03.04.2006 – 18.05.2006).

Размер выборки — 32 значения. Размер обучающей выборки — 18 значений (оптимальный для данного эксперимента).

Получены следующие результаты.

- Для треугольного вида ФП

1. Для нормированных входных данных значение критерия составило: $СКО = 0{,}057379$ (рис. 7).

2. Для ненормированных входных данных значения критериев составили: $СКО = 18{,}04$, $MAPE = 0{,}78$ (рис. 8).

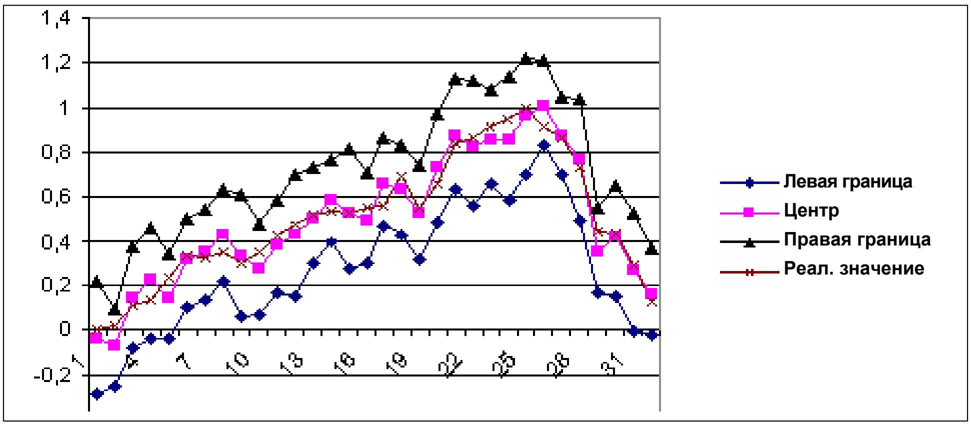

Рис. 7. Результаты эксперимента 2 для треугольной ФП и нормированных значений входных переменных

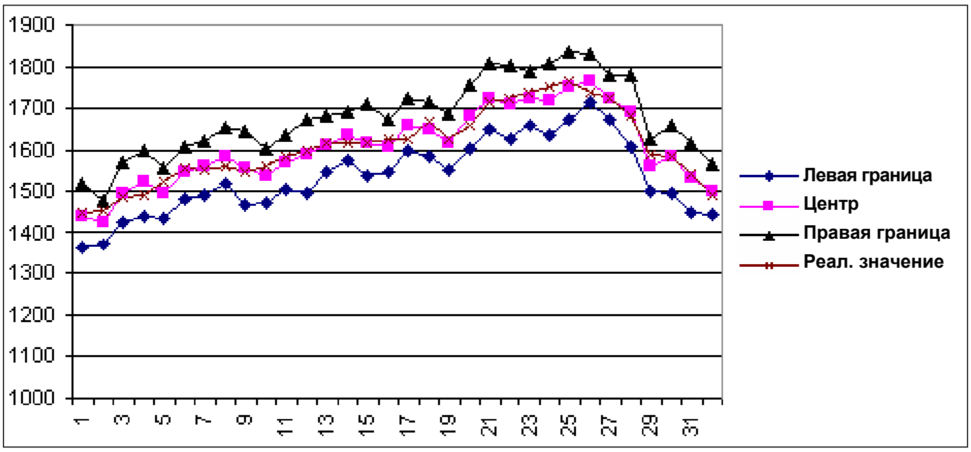

Рис. 8. Результаты эксперимента 2 для треугольной ФП и ненормированных значений входных переменных

- Для ФП Гаусса (оптимальный уровень $\alpha = 0{,}85$)

1. Для нормированных входных данных значение критерия составило: $\text{СКО} = 0{,}029582$ (рис. 9).

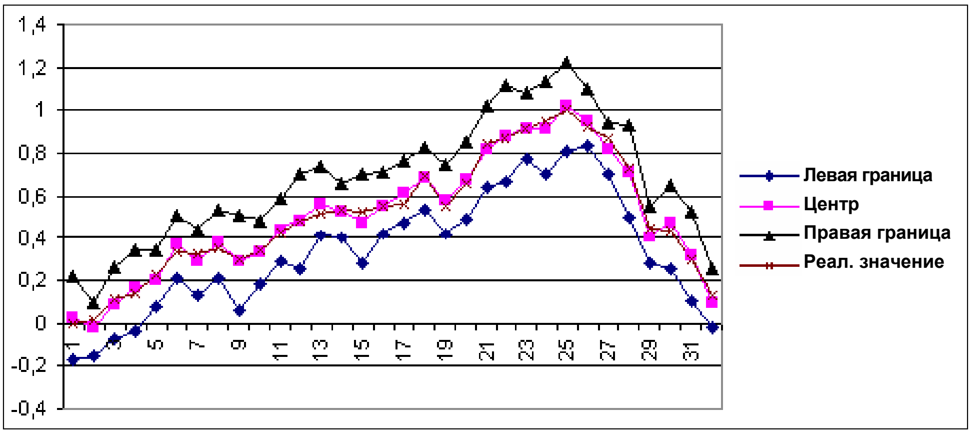

Рис. 9. Результаты эксперимента 2 для ФП Гаусса и нормированных значений входных переменных

2. Для ненормированных входных данных значения критериев составили: $СКО = 9{,}302766$, $MAPE = 0{,}37\%$ (рис. 10).

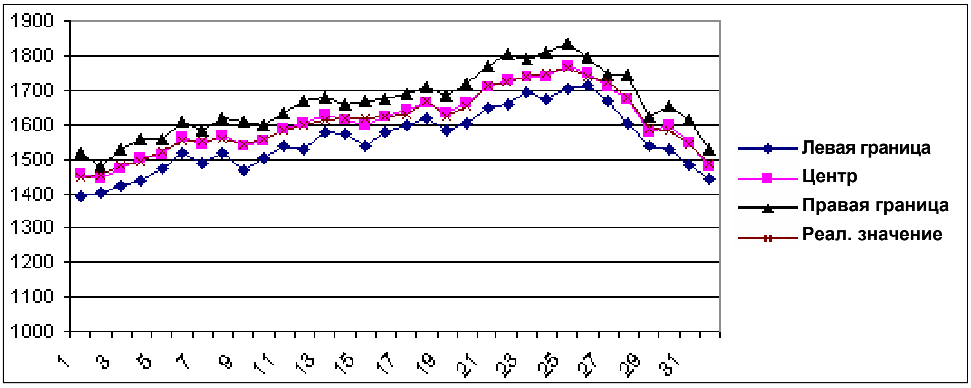

Рис. 10. Результаты эксперимента 2 для ФП Гаусса и ненормированных значений входных переменных

Как показывает эксперимент 2, прогнозирование с использованием треугольных и гауссовских ФП дает хорошие результаты, причем результаты экспериментов с ФП Гаусса лучше, чем с треугольной ФП.

Для нормированных данных

Критерии	Треугольные ФП	ФП Гаусса
СКО	0,057379	0,029582

Для ненормированных данных

Критерии	Треугольные ФП	ФП Гаусса
СКО	18,04394	9,302766
MAPE	0,78%	0,37%

ПРОГНОЗИРОВАНИЕ ИНДЕКСА РТС-2

Общая информация об индексе РТС-2

РТС-2 — это индекс акций «второго эшелона». Он является индикатором торгов на фондовой бирже РТС акциями, отнесенными ко второму эшелону по критериям ликвидности и капитализации. Расчет индекса РТС-2 выполняется в режиме и по формулам для расчета индекса РТС. Эти два индекса отличаются друг от друга списком акций, входящих в список их расчета.

Отбор акций для расчета РТС-2 выполняется Информационным комитетом РТС. В список акций не включаются наиболее ликвидные и капитализированные, а также акции, имеющие показатели ликвидности, не достаточные для корректного определения цены акций, за исключением ценных бумаг, которые входят в список наиболее ликвидных и наиболее капитализированных акций. При формировании списка, кроме показателей ликвидности и капитализации, учитывается экспертная оценка рыночных перспектив акций. Пересмотр списка для расчета РТС-2 выполняется при каждом изменении списка для расчета индекса РТС.

Изменения списка для расчета индекса РТС-2 вступают в действие 15 марта, 15 июня, 15 сентября, 15 декабря.

Результаты работы НМГУА с нечеткими входными при прогнозировании индекса РТС-2

Эксперимент 3. Прогнозирование индекса РТС-2 (цена открытия)

Использовалось пять нечетких входных переменных, которые представляют собой цены на акции второго эшелона энергетических компаний России, включенных в список для расчета индекса РТС-2:

BANE — акции ОАО «Башнефть», ENCO — акции ОАО «Сибирьтелеком», ESMO — акции ОАО «Центртелеком», IRGZ — акции ОАО «Иркутскэнерго», KUBN — акции ОАО «Южтелеком».

Исходной переменной были значения индекса РТС-2 (цена открытия) за тот же период (03.04.2006 – 18.05.2006).

Размер выборки — 32 значения.

Размер обучающей выборки — 19 значений (оптимальный размер для данного эксперимента).

Получены следующие результаты.

• Для треугольного вида ФП значение критерия составило: $СКО = 0{,}061787$ (рис. 11).

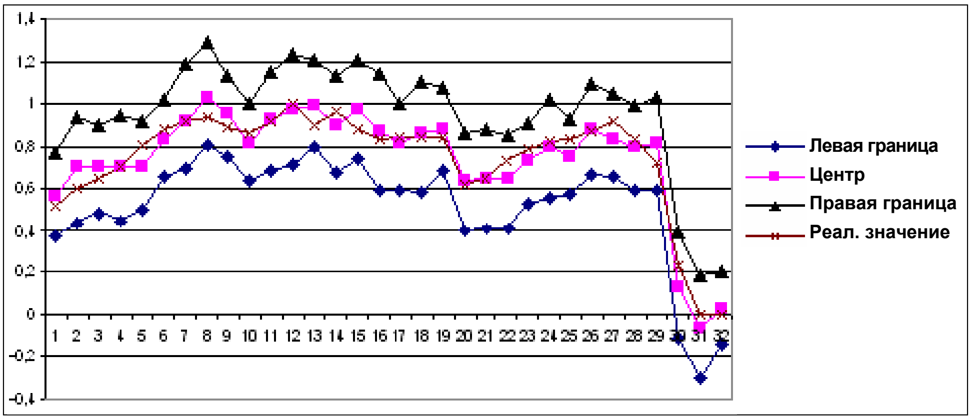

Рис. 11. Результаты эксперимента 3 для треугольной ФП и нормированных значений входных переменных

Для ненормированных входных данных значения критериев составили: $СКО = 6{,}407928$, $MAPE = 0{,}24\,\%$ (рис. 12).

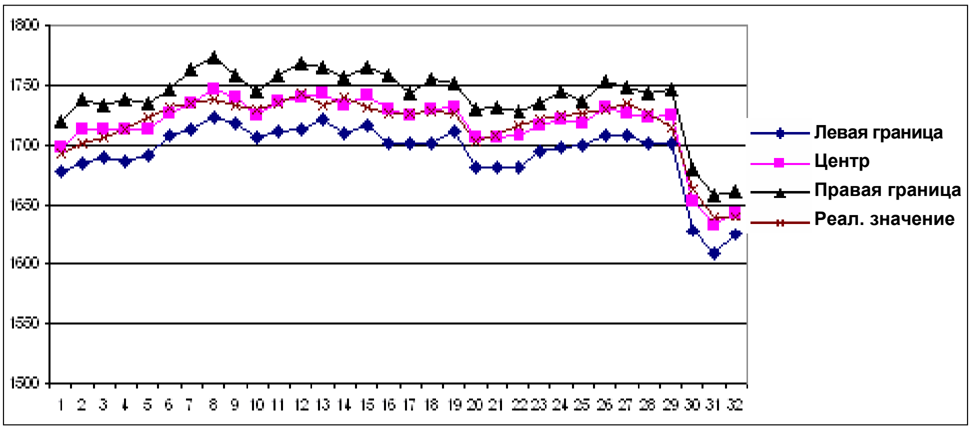

Рис. 12. Результаты эксперимента 3 для треугольной ФП и ненормированных значений входных переменных

- Для ФП Гаусса (оптимальный уровень $\alpha = 0{,}85$)

1. Для нормированных входных данных значение критерия составило: СКО $= 0{,}033097$ (рис. 13).

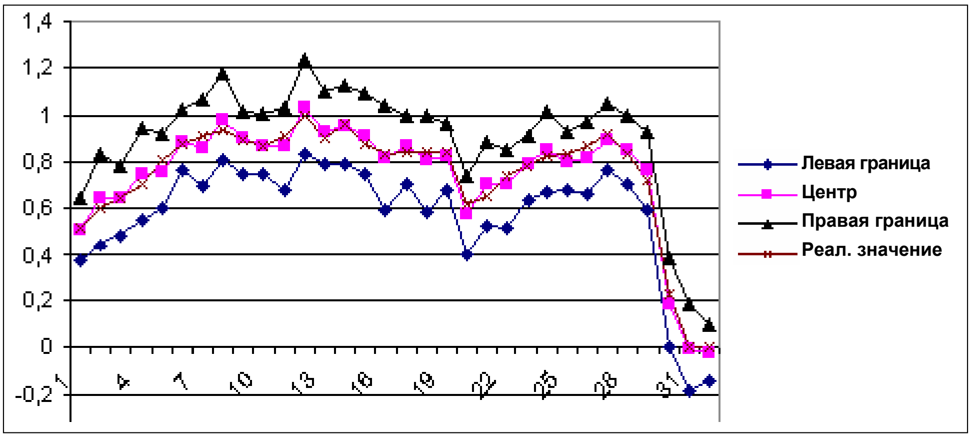

Рис. 13. Результаты эксперимента 3 для ФП Гаусса и нормированных значений входных переменных

2. Для ненормированных входных данных значения критериев составили: СКО $= 3{,}432511$, MAPE $= 0{,}13\%$ (рис. 14).

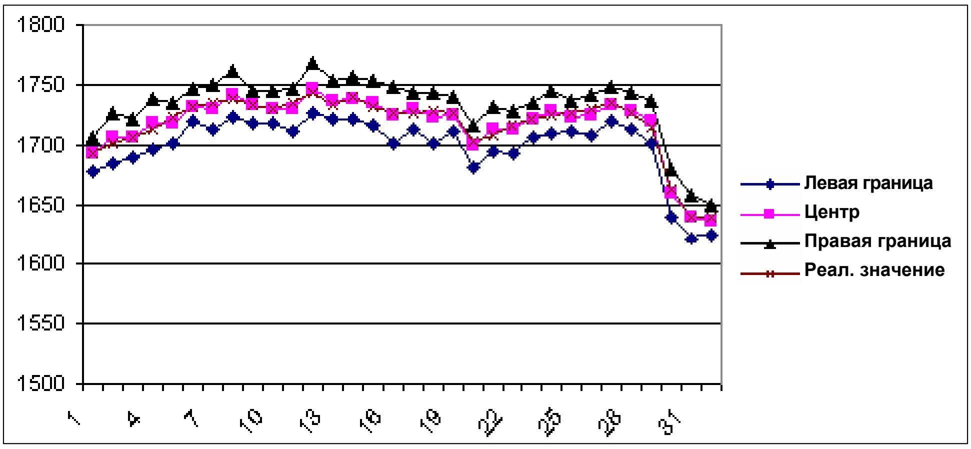

Рис. 14. Результаты эксперимента 3 для ФП Гаусса и ненормированных значений входных переменных

Как показывает эксперимент 3, прогнозирование с использованием треугольных и гауссовских ФП дает хорошие результаты. Результаты экспериментов с ФП Гаусса оказываются лучше, чем результаты с треугольной ФП.

Для нормированных данных

Критерии	Треугольные ФП	ФП Гаусса
СКО	0,061787	0,033097

Для ненормированных данных

Критерии	Треугольные ФП	ФП Гаусса
СКО	6,407928	3,432511
MAPE	0,24%	0,13%

ВЫВОДЫ

Предложен нечеткий метод группового учета аргументов с нечеткими входными данными. Он является дальнейшим развитием нечеткого метода группового учета аргументов. Метод позволяет строить зависимость между экономическими показателями, заданными нечетко, и выходным экономическим показателем. В результате работы метода можно получить доверительный интервал для исходной переменной, что является нагляднее, чем получение точечной оценки. Достоинством метода является также то, что не требуется задаваться видом прогнозирующей модели — алгоритм строит модель автоматически.

Разработана программа, реализующая предложенный метод и проведены ее экспериментальные исследования в задачах прогнозирования финансовых индексов и курсов акций на фондовом рынке (РТС). В процессе экспериментов варьировался вид функций принадлежности нечетких коэффициентов: треугольный и гауссовский.

Как следует из проведенных экспериментов, применение как треугольной, так и гауссовской ФП позволяет получить хорошие результаты. В случае ФП Гаусса они всегда лучше, чем в треугольных ФП.

Как показали исследования, оптимальный уровень α для гауссовской ФП составляет 0,8…0,9, оптимальный размер обучающей выборки — 53…60% общего размера выборки.

Полученные результаты позволяют сделать вывод о целесообразности применения разработанного метода для моделирования финансовых показателей на фондовом рынке в условиях неполноты и недостоверности исходной информации.

ЛИТЕРАТУРА

1. *Zaychenko Yu*. The fuzzy group method of data handling and its application for economical processes forecasting // Scientific Inquiry. — 2006. — **7**, № 10. — P.83–98.
2. *Исследование* нечеткого метода группового учета аргументов с различными видами нечетких описаний / Ю.П. Зайченко, И.О. Заяц, А.В. Камоцкий, Е.А. Павлюк // УСиМ. — 2003. — № 2. — С. 56–67.
3. *Зайченко Ю.П., Заєць І.О.* Порівняльний аналіз прогнозуючих моделей, побудованих за допомогою чіткого та нечіткого алгоритмів МГУА з використанням різних алгоритмів генерації нечітких прогнозуючих моделей // Матеріали міжнар. семінару з індуктивного моделювання. — Київ: НАН України, МНН інформ. технол. та систем, 2005. — С. 158–165.
4. *Зайченко Ю.П., Заєць І.О.* Порівняльний аналіз чіткого та нечіткого алгоритмів МГУА з використанням різних методів покрокової адаптації коефіцієнтів // Матеріали VII міжнар. наук.-техн. конф. «Системний аналіз та інформаційні технології». — Київ: Вид. ІПСА, 2005. — С. 121.

Поступила 29.12.2006