Досліджені нечіткі методи регресії з використанням декількох критеріїв оптимальності, визначені труднощі із узгодженням критеріїв максимального сумісництва даних і мінімальної нечіткості моделі. Розроблено критерій якості підбору коефіцієнтів регресії, які об'єднують два критерії. Показано використання розробленого критерію у задачі прогнозу електроспоживання підприємством. Приведена оцінка точності прогнозу за відносною середньомодульною похибкою

Ключеві слова: прогноз електроспоживання, нечіткий регресійний аналіз, критерії оцінки якості регресії, нечітка множина

Исследованы нечеткие методы регрессии с использованием нескольких критериев оптимальности, определены сложности в согласовании критериев максимальной совместимости данных и минимальной нечеткости модели. Разработан критерий качества подбора коэффициентов регрессии, объединяющий указанные два критерия. Показано применение разработанного критерия к задаче прогноза электропотребления предприятием. Приведена оценка точности прогноза по относительной среднемодульной погрешности

Ключевые слова: прогноз электропотребления, нечеткий регрессионный анализ, критерий оценки качества регрессии, нечеткое множество

УДК 621.311

DOI: 10.15587/1729-4061.2014.27664

РАЗРАБОТКА КРИТЕРИЯ КАЧЕСТВА ПОДБОРА КОЭФФИЦИЕНТОВ РЕГРЕССИИ В ЗАДАЧАХ ПРОГНОЗИРОВАНИЯ ЭЛЕКТРОПОТРЕБЛЕНИЯ

С. А. Тимчук
Кандидат технических наук, доцент
Кафедра автоматизации и компьютерно-интегрированных технологий
Харьковский национальный технический университет сельского хозяйства
ул. Энгельса, 19, г. Харьков, Украина, 61052
E-mail: stym@i.ua

И. А. Катюха
Аспирант
Кафедра теоретической и общей электротехники
Таврический государственный агротехнологический университет
ул. Б. Хмельницкого 17, г. Мелитополь, Украина, 72315
E-mail: igorkat@mail.ru

1. Введение

В условиях современного энергорынка для крупных предприятий важное значение имеет создание системы прогнозирования потребления электрической энергии, позволяющей минимизировать отклонения потребляемой от заявленной на рынке за сутки вперед мощности.

Известно, что электропотребление потребителей носит случайный характер и зависит от множества меняющихся факторов (например, нагрузка потребителей, условия окружающей среды и др.), для решения указанных задач применяются методы, основанные на детерминированных и вероятностно-статистических методах. Самым распространенным методом на сегодняшний день является регрессионный анализ.

Классический регрессионный анализ, основанный на критерии наименьших квадратов, позволяет оценивать и прогнозировать электропотребление в зависимости от сколь угодно большого числа воздействующих факторов. Однако для построения регрессионной модели необходимо иметь большой объем исходной статистической информации о воздействующих факторах и электропотреблении.

В некоторых случаях по разным причинам информации об электропотреблении может быть недостаточно, или она может иметь интервальный или нечеткий характер, то есть представляться в виде интервальных значений или функций принадлежности. В таком случае применение традиционных методов оценки и прогнозирования электропотребления становится весьма затруднительным, что дает возможность использования подходов, предназначенных для решения описываемых задач в условиях неопределенности. Одним из таких методов является предлагаемый авторами нечеткий регрессионный анализ с модифицированным критерием, позволяющий решать задачи прогнозирования электропотребления предприятием в условиях неопределенности.

2. Постановка проблемы

Данная задача связана с разрешением проблемы неопределенности исходной информации. Одним из способов разрешения неопределенности является построение прогнозных моделей с использованием нечеткого регрессионного анализа, который в отличие от обычного регрессионного анализа, основанного на теории вероятностей, основан на теории возможностей и теории нечетких множеств [1, 2].

3. Анализ литературных источников по теме исследования

К настоящему времени разработаны различные нечеткие методы регрессии с использованием нескольких критериев оптимальности для определения нечеткого центра и разбросов.

Во-первых, это нечеткий регрессионный анализ по критерию минимальной нечеткости. В работах [1, 2] теоретически обосновано его применение. Данный критерий не учитывает внутреннюю структуру обрабатываемых нечетких данных. Когда это несущественно, применяется нечеткая интервальная регрессия, в частности при решении прогнозных задач в энергетике [3, 4] или описания исходных данных [5, 6]. Для более точного прогноза электропотребления необходимо знать не только интервал, но и структуру нечеткого значения внутри интервала.

Во-вторых, это нечеткий регрессионный анализ, комбинированный с методом наименьших квадратов (FLSRA). Он фигурирует в двух вариантах: FLSRA по критерию минимальной нечеткости [1, 7] и FLSRA по критерию максимальной совместимости [1, 8]. Для первого варианта характерен следующий недостаток: чем выше степень достоверности, тем выше нечеткость регрессионной модели. Второй вариант основан на мере совместимости между исходными данными и приспособленной моделью. Понятие меры совместимости заключается в следующем. Пусть $\mu_A(X)$ и $\mu_B(X)$ – функции принадлежности двух нечетких множеств A и B, тогда мерой совместимости между A и B является $\gamma(A,B)$, которая может быть выражена таким образом:

$$\gamma(A,B)=\max\min\left\{\mu_A(X),\mu_B(X)\right\}. \qquad (1)$$

Данный вариант может использовать в качестве исходных данных четкие и нечеткие значения зависимой переменной. В качестве недостатка данного варианта можно отметить, во-первых, чем выше желательная совместимость между данными и моделью, тем больше нечеткий разброс, во-вторых, уравнение центральной линии отличается от обычного МНК – уравнения регрессии, в-третьих, непопулярность вследствие чрезмерной сложности вычислений.

Ряд работ посвящены методам определения коэффициентов регрессии. Они однозначно привязаны к виду регрессионных зависимостей. Проще всего находятся коэффициенты регрессии в линейных моделях [9, 10], интересен также подход к решению данной задачи для небольших наборов нечетких исходных данных [11]. Здесь используются наработки МНК, а также методы линейного программирования. Для нелинейных моделей применяются методы нелинейного программирования, в общем виде описанные в [12] и даже эволюционные алгоритмы [13]. В контексте задачи прогнозирования электропотребления могут применяться все эти методы. В каждом конкретном случае все определяется сложностью прогнозной модели.

Анализируя в целом наработки в области нечеткого регрессионного анализа можно отметить, что основная сложность заключается в согласовании двух противоречивых критериев: максимальной совместимости данных и модели и минимальной нечеткости модели.

4. Цель и задачи исследования

Целью работы является совершенствование системы прогнозирования электропотребления промышленного предприятия в условиях неопределенности исходной информации на основе нечеткого регрессионного анализа.

Соответственно, задачами исследования, необходимыми для достижения поставленной цели, являются: разработка комбинированного критерия, гармонизирующего указанные два критерия; определение и обоснование его связи с традиционными критериями оценки точности прогнозных моделей; выбор алгоритма определения нечетких коэффициентов регрессии.

5. Разработка критерия качества подбора коэффициентов регрессии

В общем случае предпочтителен подход, когда подбор коэффициентов регрессии осуществляется по критерию максимальной совместимости, поскольку он отражает качество описания напрямую, а не косвенно, как это принято в МНК – по минимуму суммы квадратов невязок. Мера совместимости (1) между нечеткими множествами удобна для случаев симметричных нечетких чисел и не учитывает особенности их внутренней структуры. Кроме того она неоднозначна.

Между тем, меру совместимости нечетких множеств можно довольно точно оценить по их пересечению. Это положение легло в основу для разработки критерия качества подбора коэффициентов регрессии.

Параметр y, заданный в виде треугольного нечеткого числа, определяется кортежем из трех чисел $y=<y_{cp}, y_L, y_R>|_x$.

Степень близости треугольных нечетких чисел можно оценить по величине их пересечения $S=y_1\cap y$, где $y_1=< y_{1cp}, y_{1L}, y_{1R}>|_x$.– нечеткая оценка. Пересечение нечетких чисел в общем случае имеет функцию принадлежности, отличную от треугольной (на рис. 1 выделено жирным) и иметь высоту $h\neq1$. Численно пересечение треугольных нечетких чисел отражает площадь фигуры под функцией принадлежности пересечения $S_\cap = S_\Delta \cap S_{\Delta 1}$.

Тогда если имеется n значений параметра y при разных значениях x, то степень близости расчетных значений к исходным данным будет отражать величина

$$S_\cap = \sum_{i=1}^{n} S_{\cap i}\ . \qquad (2)$$

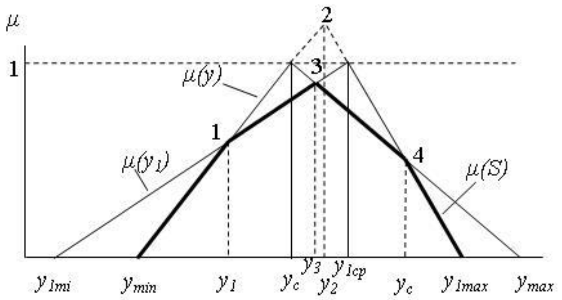

Рис. 1. Пересечение треугольных нечетких чисел

Однако величина $S_\cap$ не может служить единственным критерием для подбора коэффициентов регрессии, поскольку максимум $S_\cap$ соответствует максимуму $S_{1\Delta}$, что приведет к неоправданному увеличению степени нечеткости регрессии.

Поэтому для подбора коэффициентов регрессии предлагается использовать и критерий минимальной нечеткости. Во избежание решения задачи в условиях двухкритериальности, предлагается использовать алгебраическую свертку данных двух критериев в виде:

$$S=\sum_{i=1}^{n}\left(S_{\Delta 1i}-S_{\cap i}\right)+\sum_{i=1}^{n}\left(S_{\Delta i}-S_{\cap i}\right)\to \min . \tag{3}$$

Первая сумма в данном случае характеризует степень нечеткости, а вторая – степень совместимости. Разности в (3) введены для обеспечения процесса минимизации свертки критериев. Полученная свертка является обобщенным критерием подбора коэффициентов регрессии.

Площадь фигуры под функцией принадлежности пересечения треугольных нечетких чисел определяется для различных вариантов пересечений по следующим зависимостям, которые достаточно просто выводятся в результате геометрических построений:

1) $y_{min}<y_{1min};y_{cp}<y_{1cp};y_{max}<y_{1max}\Rightarrow S_\cap=\dfrac{(y_{max}-y_{1min})\cdot h_3}{2}.$

2) $y_{min}<y_{1min};\ y_{cp}<y_{1cp};\ y_{max}>y_{1max}\Rightarrow$
$$S_\cap=\frac{\left(y_3-y_{1min}\right)\cdot h_3}{2}+\frac{\left(y_4-y_3\right)\cdot\left(h_3+h_4\right)}{2}+\frac{\left(y_{1max}-y_4\right)\cdot h_4}{2}$$

3) $y_{min}>y_{1min};y_{cp}>y_{1cp};y_{max}<y_{1max}\Rightarrow$
$$S_\cap=\frac{(y_2-y_{min})\cdot h_2}{2}+\frac{(y_4-y_2)\cdot(h_2+h_4)}{2}+\frac{(y_{max}-y_4)\cdot h_4}{2}.$$

4) $y_{min}<y_{1min};y_{cp}>y_{1cp};y_{max}<y_{1max}\Rightarrow$
$$S_\cap=\frac{(y_1-y_{1min})\cdot h_1}{2}+\frac{(y_2-y_1)\cdot(h_1+h_2)}{2}+$$
$$+\frac{(y_4-y_2)\cdot(h_2+h_4)}{2}+\frac{(y_{max}-y_4)\cdot h_4}{2}. \tag{4}$$

5) $y_{min}<y_{1min};y_{cp}>y_{1cp};y_{max}>y_{1max}\Rightarrow$
$$S_\cap=\frac{(y_1-y_{min})\cdot h_1}{2}+\frac{(y_2-y_1)\cdot(h_1+h_2)}{2}+\frac{(y_{1max}-y_2)\cdot h_2}{2}.$$

6) $y_{min}>y_{1min};y_{cp}>y_{1cp};y_{max}>y_{1max}\Rightarrow S_\cap=\dfrac{\left(y_{1max}-y_{min}\right)\cdot h_2}{2}.$

7) $y_{min}>y_{1min};y_{cp}<y_{1cp};y_{max}>y_{1max}\Rightarrow$
$$S_\cap=\frac{(y_1-y_{min})\cdot h_1}{2}+\frac{(y_3-y_1)\cdot(h_1+h_3)}{2}+\frac{(y_4-y_3)\cdot(h_3+h_4)}{2}+$$
$$+\frac{(y_{1max}-y_4)\cdot h_4}{2}.$$

8) $y_{min}>y_{1min};y_{cp}<y_{1cp};y_{max}<y_{1max}\Rightarrow$
$$S_\cap=\frac{(y_1-y_{min})\cdot h_1}{2}+\frac{(y_3-y_1)\cdot(h_1+h_3)}{2}+\frac{(y_{max}-y_3)\cdot h_3}{2}.$$

Входящие в зависимости (4) переменные вычисляются по следующим формулам:

$$\begin{aligned}
&h_1=(y_{min}-y_{1min})\big/\left[(y_{1cp}-y_{1min})-(y_{cp}-y_{min})\right],\\
&h_2=(y_{1max}-y_{min})\big/\left[(y_{1max}-y_{1cp})+(y_{cp}-y_{min})\right],\\
&h_3=(y_{max}-y_{1min})\big/\left[(y_{max}-y_{cp})+(y_{1cp}-y_{1min})\right],\\
&h_4=(y_{max}-y_{1max})\big/\left[(y_{max}-y_{cp})+(y_{1max}-y_{1cp})\right].\\
&y_1=(y_{min}y_{1cp}-y_{1min}y_{cp})\big/\left[(y_{1cp}-y_{1min})-(y_{cp}-y_{min})\right],\\
&y_2=(y_{1max}y_{cp}-y_{min}y_{1cp})\big/\left[(y_{1max}-y_{1cp})+(y_{cp}-y_{min})\right],\\
&y_3=(y_{max}y_{1cp}-y_{1min}y_{cp})\big/\left[(y_{max}-y_{cp})+(y_{1cp}-y_{1min})\right],\\
&y_4=(y_{max}y_{1cp}-y_{1max}y_{cp})\big/\left[(y_{max}-y_{cp})+(y_{1max}-y_{1cp})\right],
\end{aligned} \tag{5}$$

где $y_{min}=y_{cp}-y_L, y_{max}=y_{cp}+y_R, y_{1min}=y_{1cp}-y_{1L}, y_{1max}=y_{1cp}+y_{1R}$.

Разработанный вариант критерия ввиду своей универсальности может быть применен к задаче прогноза на основе статистической информации, представленной в виде последовательности замеров, полученных с помощью АСКУЭ. В данном случае универсальность данного аппарата вытекает из того положения, что однозначные результаты измерений, представленные в виде детерминированного временного ряда, являются частным случаем нечеткого представления данных. То есть для нечеткого регрессионного анализа они являются синглтонами.

Параметр y, заданный в виде синглтона может быть представлен, как треугольное нечеткое число у котрого $y_{min}=y_{cp}=y_{max}$.

Тогда пересечение нечеткого треугольного числа y_1 (оценки) с синглтоном y можно оценить не по площади пересечения фигур под функциями принадлежности, а по значению функции принадлежности оценки $\mu_{y_1}(y)\big|_x$ (на рис. 2 выделено жирным).

Таким образом, если имеется n значений параметра y при разных значениях x, то степень близости оценок и исходных данных будет отражать величина

$$\mu_\cap=\sum_{i=1}^{n}\mu_{y1}\left(y_i\right). \tag{6}$$

Тогда выражение (3) примет вид:

$$S=\sum_{i=1}^{n}\left(y_{1\max i}-y_{1\min i}\right)+\sum_{i=1}^{n}\left(1-\mu_{y1}\left(y_i\right)\right)\rightarrow \min . \quad (7)$$

В результате коэффициенты регрессии определяются в виде треугольных нечетких чисел вида: $a_i=<a_{icp}, a_{iL}, a_{iR}>$.

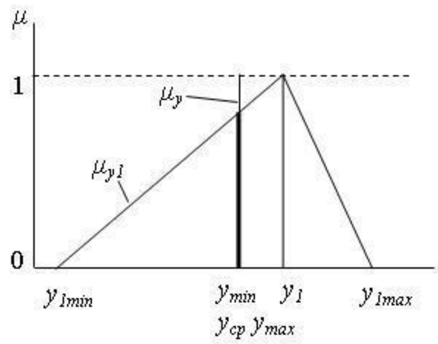

Рис. 2. Пересечение треугольного нечеткого числа и синглтона

Поскольку характер целевых функций (3), (7) заранее не известен, то для осуществления процесса поиска коэффициентов регрессии можно применить алгоритм поиска глобального оптимума, например, пространственной сетки с переменным шагом, как наиболее простой и абсолютно сходящийся с заданной точностью.

6. Оценка точности нечетких регрессионных моделей

Как и в обычном регрессионном анализе, после определения уравнение регрессии необходимо оценить соответствие регрессионной модели исследуемому объекту или процессу. Для этого производят статистический анализ нечеткого уравнения регрессии, то есть оценивают надежность нечеткого регрессионного анализа, иначе говоря, точность предсказываемых оценок.

При построении прогнозных регрессионных моделей традиционно точность прогноза оценивается по относительной среднемодульной погрешности (MAPE). Для случая нечеткой оценки данную погрешность имеет смысл вычислять для модальных значений

$$MAPE=\frac{1}{n}\sum_{i=1}^{n}\frac{\left|y_{ci}-y_{1ci}\right|}{y_{ci}}, \%. \quad (8)$$

Установим взаимосвязь степени совместимости (6) и среднемодульной погрешности (8).

Используя приведенные на рис. 3 построения, получим

$$\frac{y_i-y_{1ci}}{y_{1ci}-y_{1i\min}}=\frac{1-\mu_{yi}}{1}.$$

Отсюда, принимая во внимание различное положение y относительно y1c, получим

$$\left|y_i-y_{1ci}\right|=\begin{cases}\left(1-\mu_{yi}\right)y_{Li}, \text{ при } y_i<y_{1ci},\\ \left(1-\mu_{yi}\right)y_{Ri}, \text{ при } y_i>y_{1ci},\end{cases}$$

тогда

$$MAPE=\frac{1}{n}\sum_{i=1}^{n}\left(1-\mu_{yi}\right)y_{Ti}/y_i, \quad (9)$$

где $y_{Ti}=y_{Li}$ при $y_i<y_{ci}$ и $y_{Ti}=y_{Ri}$ при $y_i>y_{ci}$.

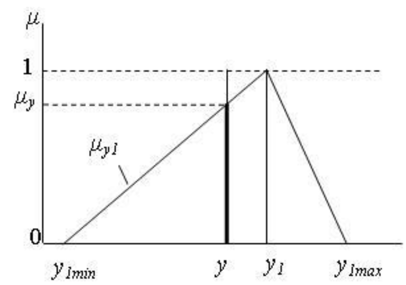

Рис. 3. Определение точности описания

Таким образом, имея степень совместимости, полученную на основе нечеткой прогнозной модели, можно с помощью (9) привести ее в соответствие с общепринятой нормой оценки погрешности.

7. Алгоритм поиска коэффициентов регрессии

В отличие от классического в нечетком регрессионном анализе определить коэффициенты регрессии аналитически в общем случае невозможно. Объясняется это особенностями построения критерия точности обработки данных. Также влияет вид самой зависимости, которая используется для построения регрессионной модели.

Следует отметить, что традиционно применяемые полиномиальные регрессионные модели обладают слабыми экстраполяционными возможностями. Поэтому прогнозные модели могут быть построены на основе более сложных функций. Что касается функциональной зависимости (7), то она может быть любой, в том числе и нелинейной относительно коэффициентов регрессии.

Таким образом, в общем случае для поиска коэффициентов регрессионной прогнозной модели на наш взгляд следует использовать методы нелинейного программирования.

Предпочтение следует отдавать абсолютно сходящимся методам поиска глобального оптимума. К таковым относится метод пространственной сетки с переменным шагом, который и был выбран в данной работе. Суть его можно изложить в форме следующего алгоритма.

1. Определение начальных значений коэффициентов регрессии.

2. Задание левого и правого коэффициента нечеткости каждого коэффициента регрессии.

3. Задание точности определения коэффициентов регрессии.

4. Задание шага изменения модальных значений и коэффициентов нечеткости коэффициентов регрессии.

5. Перебор всех возможных сочетаний модальных значений коэффициентов регрессии и значений коэффициентов их нечеткости. При этом в качестве рабочего оптимума запоминается такое сочетание, при котором достигается минимум целевой функции (7).

6. Если шаг изменения параметров удовлетворяет принятой точности, то процесс останавливается. В противном случае пространство поиска сужается до размеров шага сетки относительно найденного рабочего оптимума. Шаг изменения варьируемых параметров уменьшается с заданным модулем и процесс продолжается с п.5.

В результате итерационного процесса определяется каждый коэффициент регрессионной модели с заданной точностью. При этом целевая функция (7) достигает своего минимума.

Время поиска решения зависит в числе прочего от начального размера пространства поиска, шага сетки и модуля изменения шага по мере приближения к оптимуму.

9. Выводы

В результате решения поставленных задач можно сделать следующие выводы.

Предложен вариант критерия оптимальности, позволяющий определять коэффициенты нечеткой регрессии, исходя из паритетного учета двух противоречивых критериев: максимальной совместимости данных и модели и минимальной нечеткости модели. Разработанный вариант критерия ввиду своей универсальности может быть применен в различных задачах прогнозирования, в частности для определения будущего электропотребления предприятием. Данные для составления прогноза могут быть заданы интервально, нечетко или детерминировано в виде последовательных замеров автоматическими системами контроля и учета электроресурсов.

Математически установлена однозначная взаимосвязь предложенного критерия оптимальности с общепринятой оценкой качества прогноза на основе относительной среднемодульной погрешности.

Предложен алгоритм определения нечетких коэффициентов регрессии на основе метода пространственной сетки для случая нелинейных прогнозных моделей.

В целом полученные результаты позволяют повысить информативность системы прогнозирования электропотребления предприятия при неопределенности исходной информации.

Литература

1. Yun-Hsi, O. Chang Fuzzy regression methods - a comparative assessment. [Text] / O. Chang Yun-Hsi, M. Ayyub Bilal // Fuzzy Sets and Systems. – 2001. – Vol. 119, Issue 2. – P. 187–203. doi: 10.1016/S0165-0114(99)00091-3
2. Таранцев, А. А. Принципы построения регрессионных моделей при исходных данных с нечетким описанием. [Текст] / А. А. Таранцев // Автоматика и телемеханика. – 1997. – № 11. – С. 215–220.
3. Манусов, В. З. Методы оценивания потерь электроэнергии в условиях неопределенности. [Текст] / В. З. Манусов, А. В. Могиленко // Электричество. – 2003. – № 3. – С. 2–8.
4. Tau Lee, H. Fuzzy regression model with fuzzy input and output data for manpower forecasting [Text] / H. Tau Lee, S. Hua Chen // Fuzzy Sets and Systems. – 2001 – Vol. 119, Issue 2. – P. 205–213. doi: 10.1016/S0165-0114(98)00382-0
5. Yu, J.-R. General fuzzy piecewise regression analysis with automatic change-point detection [Text] / J.-R. Yu, G.-H. Tzeng, H.-L. Li // Fuzzy Sets and Systems. – 2001. – Vol. 119, Issue 2. – P. 247–257. doi: 10.1016/S0165-0114(98)00384-4
6. Chen, Y.-S. Outliers detection and confidence interval modification in fuzzy regression [Text] / Y.-S. Chen // Fuzzy Sets and Systems. – 2001. – Vol. 119, Issue 2. – P. 259–272. doi: 10.1016/S0165-0114(99)00049-4
7. Wang, H.-F. Insight of a fuzzy regression model [Text] / H.-F. Wang, R.-C. Tsaur // Fuzzy Sets and Systems. – 2000. – Vol. 112, Issue 3. – P. 355–369. doi: 10.1016/S0165-0114(97)00375-8
8. Celmins, A. Least squares model fitting to fuzzy vector data [Text] / A. Celmins // Fuzzy Sets and Systems – 1987. – Vol. 22. – P. 245–269. doi: 10.1016/0165-0114(87)90070-4
9. Yen K. K. A linear regression model using triangular fuzzy number coefficients [Text] / K. K. Yen, S. Ghoshray, G. Roig // Fuzzy Sets and Systems. – 1999. – Vol. 106. – P. 167–177. doi: 10.1016/S0165-0114(97)00269-8
10. Soliman, S. A. Application of -fc fuzzy linear regression algorithm to power system voltage measurements [Text] / S. A. Soliman, Abdel Rahman H. Mansour, M.E. El-Hawary // Electric Power Systems Research. – 2000. – Vol. 42, Issue 3. – P. 195–200. doi: 10.1016/s0378-7796(96)01205-9
11. Seraya, O. V. Linear regression analysis of a small sample of fuzzy input data [Text] / O. V. Seraya, D. A. Demin // Journal of Automation and Information Sciences. – 2012. – Vol. 44, Issue 7. – P. 34–48. doi: 10.1615/jautomatinfscien.v44.i7.40
12. Раскин, Л. Г. Нечеткая математика [Текст]: моногр. / Л. Г. Раскин, О. В. Серая. – Харьков: Парус, 2008. – 352 с.
13. Buckley, J. J. Linear and non-linear fuzzy regression: Evolutionary algorithm solutions. [Text] / J. J. Buckley, T. Feuring // Fuzzy Sets and Systems. – 2000. – Vol. 112, Issue 3. – P. 381–394. doi: 10.1016/S0165-0114(98)00154-7