https://orcid.org/0000-0003-3901-0719
https://orcid.org/0000-0001-7658-2850
https://orcid.org/0000-0002-4699-7868

УДК 519.712.3

М.Дж. МАРДАНОВ*, Р.Р. РЗАЕВ, П.Э. АЛИЗАДЕ*****

ОБ ОДНОМ ПОДХОДЕ К ФАЗЗИФИКАЦИИ ДАННЫХ НА ПРИМЕРЕ ВРЕМЕННОГО РЯДА ИНДЕКСА ДОУ-ДЖОНСА

*Институт математики и механики НАН Азербайджана, г. Баку, Азербайджан
**Институт систем управления НАН Азербайджана, г. Баку, Азербайджан
***Бакинский государственный университет, г. Баку, Азербайджан

Анотація. *Природа слабоструктурованих систем обумовлена використанням експертних оцінок, притаманна невизначеність яких відноситься до класу нечіткості. На відміну від стохастичної невизначеності нечіткість ускладнює або навіть виключає застосування статистичних методів і моделей, але може бути використана для прийняття предметно-орієнтованих рішень на основі наближених міркувань людини. Формалізація інтелектуальних операцій, які моделюють нечіткі висловлювання людини про стан і поведінку слабоструктурованих систем і/або складних явищ, сьогодні утворює самостійний напрям науково-прикладних досліджень, одним з яких є нечітке моделювання тимчасових (або динамічних) рядів. Цей напрям включає комплекс завдань, методологія вирішення яких спирається на теорію нечітких множин, нечіткої логіки і нечітких моделей (або систем нечіткого виводу). Початковою процедурою нечіткого моделювання часових рядів є фазифікація історичних даних, отриманих шляхом спостережень на основі «м'яких вимірів» за поведінкою динамічної системи за певний проміжок часу. У статті пропонується нове правило фазифікації таких даних, яке апробовано на показниках індустріального індексу Доу-Джонса, що встановлюється за результатами щоденних торгів на фондовій біржі США шляхом звичайного арифметичного усереднення складових показників. Пропонована у статті процедура фазифікації реалізується через систему нечіткого виводу, яка забезпечує знаходження значень функцій приналежності відповідних нечітких підмножин дискретного універсуму, що покриває всю сукупність показників індексу Доу-Джонса за більш ніж річний період.*

Ключові слова: *індекс Доу-Джонса, історична дана, нечіткий часовий ряд, функція приналежності, нечіткий висновок.*

Аннотация. *Природа слабоструктурированных систем обусловлена использованием экспертных оценок, присущая неопределённость которых относится к классу нечёткости. В отличие от стохастической неопределённости нечёткость затрудняет или даже исключает применение статистических методов и моделей, но может быть использована для принятия предметно-ориентированных решений на основе приближенных рассуждений человека. Формализация интеллектуальных операций, моделирующих нечёткие высказывания человека о состоянии и поведении слабоструктурированных систем и/или сложных явлений, сегодня образует самостоятельное направление научно-прикладных исследований, одним из которых является нечёткое моделирование временных (или динамических) рядов. Это направление включает комплекс задач, методология решения которых опирается на теорию нечётких множеств, нечёткой логики и нечётких моделей (или систем нечёткого вывода). Начальной процедурой нечёткого моделирования временных рядов является фаззификация исторических данных, полученных путём наблюдений на основе «мягких измерений» за поведением динамической системы за определённый промежуток времени. В статье предлагается новое правило фаззификации таких данных, которое апробировано на показателях индустриального индекса Доу-Джонса, устанавливаемого по результатам ежедневных торгов на фондовой бирже США путём обычного арифметического усреднения составляющих показателей. Предлагаемая в статье процедура фаззификации реализуется через систему нечёт-*

кого вывода, которая обеспечивает нахождение значений функций принадлежности соответствующих нечётких подмножеств дискретного универсума, покрывающего всю совокупность показателей индекса Доу-Джонса за более чем годичный период.
Ключевые слова: *индекс Доу-Джонса, историческая данная, нечёткий временной ряд, функция принадлежности, нечёткий вывод.*

Abstract. *The nature of weakly structured systems is specified by using of expert estimates, the inherent uncertainty of which belongs to the fuzzy class. Unlike stochastic uncertainty, fuzziness complicates or even eliminates the use of statistical methods and models, but can be used to make subject-oriented decisions based on approximate reasoning of a person. The formalization of intellectual operations simulating fuzzy human statements about the state and behavior of weakly structured systems and/or complex phenomena today forms an independent direction of scientific and applied researches, one of which is fuzzy modeling of time (or dynamic) series. This direction includes a task complex, the methodology of which is based on the theory of fuzzy sets, fuzzy logic and fuzzy models (or fuzzy inference systems). The initial procedure for fuzzy modeling of time series is fuzzification of historical data obtained by observing on the basis of "soft measurements" of the behavior of a dynamic system for a certain period of time. The paper proposes a new rule of fuzzification of such data, which is tested on the indicators of the Dow Jones Industrial Average, established by the results of daily trading on the US stock exchange by the usual arithmetic averaging of component indicators. The fuzzification procedure proposed in the given paper is implemented through a fuzzy inference system, which ensures that the membership functions of the corresponding fuzzy subsets of the discrete universe are found, covering the entire set of indicators of the Dow Jones index for more than a year.*
Keywords: *Dow Jones index, historical data, fuzzy time series, membership function, fuzzy inference.*

DOI: 10.34121/1028-9763-2020-2-3-13

1. Введение

Большинство существующих алгоритмов прогнозирования временных рядов, как правило, работают с так называемыми «чистыми» данными, то есть с данными, представленными в виде действительных чисел. Однако на практике данные наблюдения за поведением той или иной системы должны рассматриваться как слабоструктурированные, то есть такими, о которых известна их принадлежность к определенному классу. Наиболее адекватными способами представления таких данных являются интервалы вида $x \in [x_{\min}, x_{\max}]$ или, что ещё лучше, высказывания вида «*x* близко к 13», подчёркивающие локализацию наблюдений относительно тех или иных структурированных величин. Поэтому, если принять к сведению, что наблюдения за поведением динамической системы через одинаковые промежутки времени осуществляются на уровне «мягких измерений», то есть в виде термов лингвистической переменной, то формируемый при этом временной (или динамический) ряд является слабоструктурированным. Ясно, что по своей природе слабоструктурированные временные ряды в корне отличаются от обычных числовых рядов, и поэтому для них необходим соответствующий механизм моделирования и прогнозирования.

В настоящее время наиболее простым способом формализации слабоструктурированных данных являются нечёткие множества [1], которые достаточно приемлемым образом отождествляют значения (термы) лингвистических переменных через построение подходящих функций принадлежности. Поэтому слабоструктурированные временные ряды целесообразно интерпретировать как нечёткие со всеми вытекающими из этого последствиями относительно их моделирования и прогнозирования.

На протяжении последних двух десятилетий исследованию нечётких временных рядов посвящены многочисленные публикации, среди которых, прежде всего, следует отметить работы К. Сонга и Б. Чиссома [2], Н. Кумара и др. [3], С. Чена [4], К. Ченга и др. [5], Дж. Поулсена [6], Р.Р. Рзаева и др. [7]. Описанные в этих работах подходы к восстановлению временных рядов отличаются правилами фаззификации исторических данных и/или дефаззификации нечётких выходов моделей. Понятно, что насколько эти правила позволяют адекватно описывать слабоструктурированные данные временного ряда посредством нечётких множеств и, соответственно, интерпретировать полученные результаты в традиционной численной манере, зависит достоверность конечных прогнозов. В статье предлагается подход к описанию слабоструктурированных данных временного ряда, основанный на применении системы нечёткого вывода.

2. Постановка задачи

Объектом исследования является временной ряд (ВР): $\{x(t)\}(t=1\div T)$, в котором $x(t)$ в силу ряда причин является слабоструктурированной исторической данной (ИД) или, в нашем представлении, нечётким множеством (НМ) $A_j(j=1\div J)$, характеризуемым картежем [1]: $\{x(t)/\mu_{Aj}[x(t)]\}$, $\mu_{Aj}[x(t)]\to[0, 1]$. Необходимо разработать метод фаззификации исторических данных, который позволил бы более адекватно восстановить ВР в терминах нечётких множеств и тем самым повысить достоверность его прогнозирования на основе существующих нечётких моделей.

Цель статьи заключается в разработке новой процедуры фаззификации исторических данных, обеспечивающей построение более адекватной нечёткой модели на примере временного ряда индекса Доу-Джонса и, тем самым, повышающей качество его прогнозирования.

3. Фаззификация исторических данных

В контексте построения системы нечёткого вывода (СНВ) под процедурой фаззификации (или введение нечёткости) понимается процесс нахождения значений функций принадлежности (ФП) нечётких множеств, описывающих термы входных лингвистических переменных. Другими словами, путём фаззификации устанавливаются соответствия между конкретным значением отдельной входной переменной СНВ и значением ФП соответствующего ей терма входной лингвистической переменной (ЛП). Различают две группы методов построения ФП: прямые и косвенные. Прямые методы характеризуются тем, что построение ФП непосредственно осуществляется экспертами, располагающими знаниями в предметной области. Примерами прямых методов фаззификации являются непосредственные задания ФП в табличном, графическом или аналитическом видах. Косвенные методы предполагают выбор значений ФП, который производится в соответствии с предварительно сформулированными экспертами условиями. К числу таких методов относятся методы построения ФП на основе попарных сравнений релевантных статистических данных, на основе экспертных ранговых оценок и др.

Как нетрудно заметить, все методы в той или иной степени опираются на эвристические знания и поэтому отличаются существенной долей субъективизма, присущей любому экспертному суждению. Исходя из этой предпосылки, становятся очевидными важность и актуальность исследования методов построения ФП в процессе фаззификации входных данных в рамках нечёткого моделирования ВР. Для введения нечёткости, как известно, необходимо определиться с универсумом. В случае ВР основанием для этого служит покрытие диапазона исторических данных (ИД). Согласно [6], таким покрытием может служить отрезок $d=[D_{\min}-D_1; D_{\max}+D_2]$, где $D_{\min}$ и $D_{\max}$ – соответственно минимальное и максимальное значения показателя ВР; D_1 и D_2 – положительные числа, выбираемые пользователем, например, из расчёта деления отрезка d на равные интервалы u_j по числу заранее выбранных критериев оценки. В частности, под термами $C_1, C_2, \ldots, C_n$ лингвистической переменной «величина показателя ВР» понимают нечёткие подмножества дискретного универсума $U=\{u_1, u_2, \ldots, u_n\}$, отражающие качественные критерии оценки величины ИД, которые обычно задаются в виде

$$C_1=\mu_{C1}(u_1)/u_1+\mu_{C1}(u_2)/u_2+\ \ldots\ +\mu_{C1}(u_n)/u_n,$$
$$C_2=\mu_{C2}(u_1)/u_1+\mu_{C2}(u_2)/u_2+\ \ldots\ +\mu_{C2}(u_n)/u_n,$$
$$\ldots$$
$$C_n=\mu_{Cn}(u_1)/u_1+\mu_{Cn}(u_2)/u_2+\ \ldots\ +\mu_{Cn}(u_n)/u_n,$$

где $\mu_{Ci}(u_j)\in[0,\ 1]$ $(i,\ j=1\div n)$ – значения функции принадлежности интервала u_j к НМ C_i. То есть, если данная принадлежит интервалу u_j, то её нечёткой интерпретацией будет множество C_i. Поэтому в данном случае необходимо правильно идентифицировать ФП $\mu_{Ci}(u)$, восстанавливающую НМ C_i.

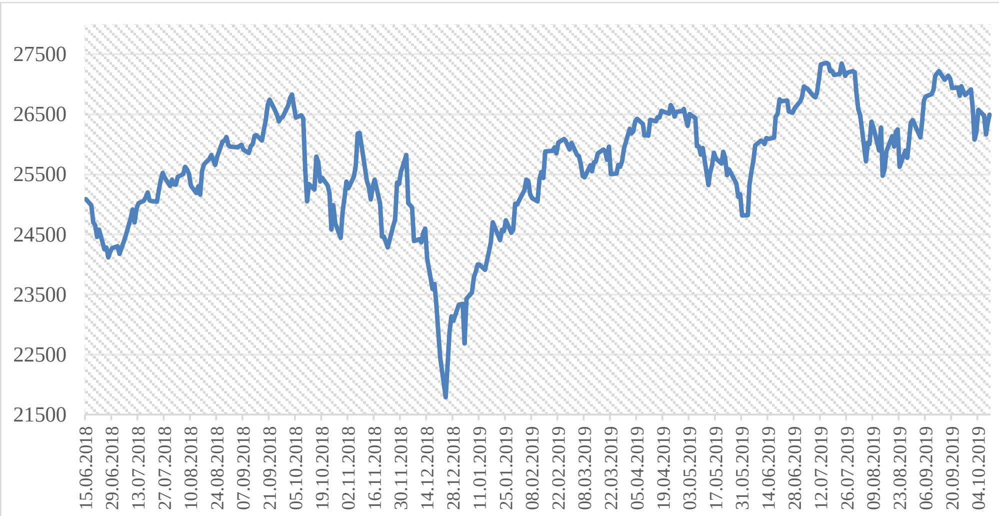

Рисунок 1 – Временной ряд DJI

Рассмотрим ВР, отражающий динамику изменения индустриального индекса Доу-Джонса (DJI) за период с 15.06.2018 г. по 10.10.2019 г. (рис. 1 и/или табл. 1) [8], где минимальным и максимальным значениями среди 333 ИД соответственно являются числа $D_{\min}=21792{,}2$ и $D_{\max}=27359{,}2$. Выбирая $D_1=21{,}2$ и $D_2=11{,}8$, получим соответствующий универсум: $U=[21771,\ 27371]$, который разобьём на восемь равных интервалов длиной в 700 единиц: u_1=[21771, 22471], u_2=[22471, 23171], u_3=[23171, 23871], u_4=[23871, 24571], u_5=[24571, 25271], u_6=[25271, 25971], u_7=[25971, 26671], u_8=[26671, 27371]. Тогда в качестве возможных значений (термов) ЛП «Величина DJI» выберем следующие термы, описываемые соответствующими нечёткими множествами $C_j\ (k=1\div 8)$:

- СЛИШКОМ НИЗКАЯ: $C_1=1/u_1+0{,}5/u_2+0/u_3+0/u_4+0/u_5+0/u_6+0/u_7+0/u_8$;
- ОЧЕНЬ НИЗКАЯ: $C_2=0{,}5/u_1+1/u_2+0{,}5/u_3+0/u_4+0/u_5+0/u_6+0/u_7+0/u_8$;
- БОЛЕЕ ЧЕМ НИЗКАЯ: $C_3=0/u_1+0{,}5/u_2+1/u_3+0{,}5/u_4+0/u_5+0/u_6+0/u_7+0/u_8$;
- НИЗКАЯ: $C_4=0/u_1+0/u_2+0{,}5/u_3+1/u_4+0{,}5/u_5+0/u_6+0/u_7+0/u_8$;
- ВЫСОКАЯ: $C_5=0/u_1+0/u_2+0/u_3+0{,}5/u_4+1/u_5+0{,}5/u_6+0/u_7+0/u_8$;
- БОЛЕЕ ЧЕМ ВЫСОКАЯ: $C_6=0/u_1+0/u_2+0/u_3+0/u_4+0{,}5/u_5+1/u_6+0{,}5/u_7+0/u_8$;
- ОЧЕНЬ ВЫСОКАЯ: $C_7=0/u_1+0/u_2+0/u_3+0/u_4+0/u_5+0{,}5/u_6+1/u_7+0{,}5/u_8$;
- СЛИШКОМ ВЫСОКАЯ: $C_8=0/u_1+0/u_2+0/u_3+0/u_4+0/u_5+0/u_6+0{,}5/u_7+1/u_8$.

В этом случае нечёткая интерпретация ИД осуществляется с учётом принадлежности интервала их локализации $u_j(j=1\div 8)$ тому или иному НМ C_j с достаточно тривиальной ФП. В результате этого подхода нечёткий аналог ВР DJI можно получить так, как это представлено в табл. 1. Рассмотрим другой подход к фаззификации ИД.

4. Фаззификация ИД на основе вербального моделирования

Опираясь на предыдущие рассуждения, за основу выбраны следующие непротиворечивые и достаточно тривиальные высказывания:

e_1: «Если ИД DJI расположена ближе к середине отрезка u_1, то её величина является слишком низкой»;

e_2: «Если ИД DJI расположена ближе к середине отрезка u_2, то её величина является очень низкой»;

e_3: «Если ИД DJI расположена ближе к середине отрезка u_3, то её величина является более чем низкой»;

e_4: «Если ИД DJI расположена ближе к середине отрезка u_4, то её величина является низкой»;

e_5: «Если ИД DJI расположена ближе к середине отрезка u_5, то её величина является высокой»;

e_6: «Если ИД DJI расположена ближе к середине отрезка u_6, то её величина является более чем высокой»;

e_7: «Если ИД DJI расположена ближе к середине отрезка u_7, то её величина является очень высокой»;

e_8: «Если ИД DJI расположена ближе к середине отрезка u_8, то её величина является слишком высокой».

Анализ этих рассуждений, как информационных фрагментов общей вербальной модели позволяет выявить одну входную характеристику в виде ЛП x=«Расположение исторической данной», значениями которой являются термы «БЛИЖЕ К СЕРЕДИНЕ ОТРЕЗКА u_k» $(k=1\div8)$, и одну выходную ЛП $y=$«Величина DJI», принимающую значения в виде термов: СЛИШКОМ НИЗКАЯ, ОЧЕНЬ НИЗКАЯ, БОЛЕЕ ЧЕМ НИЗКАЯ, НИЗКАЯ, ВЫСОКАЯ, БОЛЕЕ ЧЕМ ВЫСОКАЯ, ОЧЕНЬ ВЫСОКАЯ, СЛИШКОМ ВЫСОКАЯ. Оценку локализации ИД $x(t)$ в рамках ВР DJI по признакам её принадлежности тому или иному отрезку $u_j\,(j=1\div8)$ интерпретируем в виде нечёткого подмножества дискретного универсума, состоящего из полного набора данных DJI: $U=\{x(t)\}_{t=1\div333}$. В качестве ФП НМ выберем гауссовские функции вида $\mu(x)=\exp\{-(x_t-u_{j0})^2/\sigma^2\}$, где $x_t=x(t)$ – ИД DJI по результатам завершения торгов на фондовой бирже за t-ый день; u_{j0} – середина интервала u_j; σ^2 – параметр плотности, выбранный единым для всех случаев как 2500000 (рис. 2). Тогда, принимая во внимание середины отрезков u_j (j=1÷8): u_{10}=22121, u_{20}=22821, u_{30}=23521, u_{40}=24221, u_{50}=24921, u_{60}=25621, u_{70}=26321, u_{80}=27021, признаки локализации ИД $x(t)$ интерпретируем как

- «БЛИЗКО К 22121» как НМ: $X_1=0{,}952181/x_1+0{,}958630/x_2+$ $+0{,}964640/x_3+\ldots+0{,}000793/x_{332}+0{,}000472/x_{333}$;
- «БЛИЗКО К 22821» как НМ: $X_2=0{,}643393/x_1+0{,}656883/x_2+$ $+0{,}670320/x_3+\ldots+0{,}006941/x_{332}+0{,}004497/x_{333}$;
- «БЛИЗКО К 23521» как НМ: $X_3=0{,}293758/x_1+0{,}304145/x_2+$ $+0{,}314743/x_3+\ldots+0{,}041079/x_{332}+0{,}028958/x_{333}$;
- «БЛИЗКО К 24221» как НМ: $X_4=0{,}090627/x_1+0{,}095155/x_2+$ $+0{,}099859/x_3+\ldots+0{,}164269/x_{332}+0{,}125994/x_{333}$;
- «БЛИЗКО К 24921» как НМ: $X_5=0{,}018892/x_1+0{,}020116/x_2+$ $+0{,}021408/x_3+\ldots+0{,}443858/x_{332}+0{,}370415/x_{333}$;
- «БЛИЗКО К 25621» как НМ: $X_6=0{,}002661/x_1+0{,}002873/x_2+$ $+0{,}003101/x_3+\ldots+0{,}810382/x_{332}+0{,}735842/x_{333}$;

• «БЛИЗКО К 26321» как НМ: X_7=0,000253/x_1+0,000277/x_2+ +0,000304/x_3+…+0,999750/x_{332}+0,987728/x_{333};

• «БЛИЗКО К 27021» как НМ: X_8=0,000016/x_1+0,000018/x_2+ 0,000020/x_3+…+0,833393/x_{332}+0,895873/x_{333}.

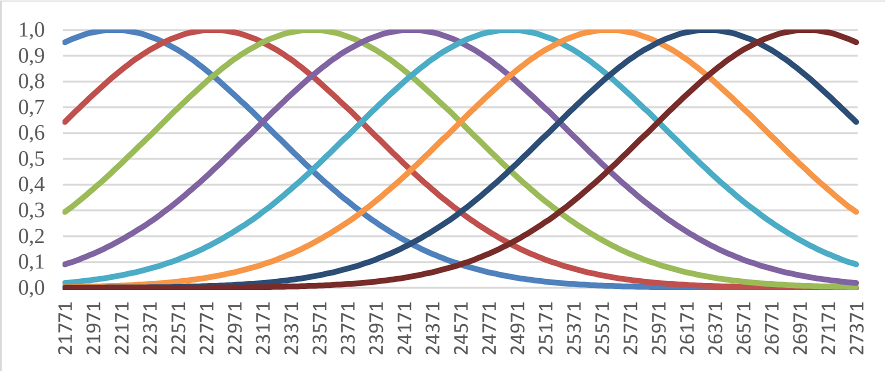

Рисунок 2 – ФП НМ, отражающих степень локализации ИД

Термы выходной ЛП «Величина DJI» опишем в виде нечётких подмножеств универсума $I = \{0;\ 0,1;\ 0,2;\ \ldots;1\}$ так, что $\forall i \in I$ имеем [7]:

• TL=СЛИШКОМ НИЗКОЕ: $\mu_{TL}(i)$=0, если i=1, и $\mu_{TL}(i)$=1, если i<1;

• VL=ОЧЕНЬ НИЗКОЕ: $\mu_{VL}(i)=(1\text{-}i)^2$;

• ML=БОЛЕЕ ЧЕМ НИЗКОЕ: $\mu_{ML}(i)=(1\text{-}i)^{1/2}$;

• L=НИЗКОЕ: $\mu_L(i)=1\text{-}i$;

• H=ВЫСОКОЕ: $\mu_H(i)=i$;

• MH=БОЛЕЕ ЧЕМ ВЫСОКОЕ: $\mu_{MH}(i)=i^{1/2}$;

• VH=ОЧЕНЬ ВЫСОКОЕ: $\mu_{VH}(i)=i^2$;

• TH=СЛИШКОМ ВЫСОКОЕ, $\mu_{TH}(i)$=1, если i=1, и $\mu_{TH}(i)$=0, если i<1.

Тогда, с учётом введённых формализмов, приведённые выше рассуждения запишем в виде следующих нечётких импликативных правил:

e_1: «Если $x=X_1$, то $y=TL$»; e_2: «Если $x=X_2$, то $y=VL$»; e_3: «Если $x=X_3$, то $y=ML$»;
e_4: «Если $x=X_4$, то $y=L$»; e_5: «Если $x=X_5$, то $y=H$»; e_6: «Если $x=X_6$, то $y=MH$»;
e_7: «Если $x=X_7$, то $y=VH$»; e_8: «Если $x=X_8$, то $y=TH$».

Для преобразования этих правил применена импликация Лукасевича [1, 7]: $\mu_W(u,\ i)=\min\{1,\ 1\text{-}\mu_X(u)+\mu_Y(i)\}$, в результате чего для каждой пары $(u,\ i)\in X\times Y$ на $X\times Y$ были получены соответствующие нечёткие отношения:

$$R_1 = \left[\begin{array}{c|ccccccccccc} & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0,0294 & 1,0000 & 0,9706 & 0,9706 & 0,9706 & 0,9706 & 0,9706 & 0,9706 & 0,9706 & 0,9706 & 0,9706 & 0,9706 \\ 0,0374 & 1,0000 & 0,9626 & 0,9626 & 0,9626 & 0,9626 & 0,9626 & 0,9626 & 0,9626 & 0,9626 & 0,9626 & 0,9626 \\ 0,0699 & 1,0000 & 0,9301 & 0,9301 & 0,9301 & 0,9301 & 0,9301 & 0,9301 & 0,9301 & 0,9301 & 0,9301 & 0,9301 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0,0008 & 1,0000 & 0,9992 & 0,9992 & 0,9992 & 0,9992 & 0,9992 & 0,9992 & 0,9992 & 0,9992 & 0,9992 & 0,9992 \\ 0,0005 & 1,0000 & 0,9995 & 0,9995 & 0,9995 & 0,9995 & 0,9995 & 0,9995 & 0,9995 & 0,9995 & 0,9995 & 0,9995 \end{array}\right],$$

$$R_2=\left[\begin{array}{c|ccccccccccc}
 & 1 & 0,81 & 0,64 & 0,49 & 0,36 & 0,25 & 0,16 & 0,09 & 0,04 & 0,01 & 0 \\ \hline
0,1274 & 1,0000 & 1,0000 & 1,0000 & 1,0000 & 1,0000 & 1,0000 & 1,0000 & 0,9626 & 0,9126 & 0,8826 & 0,8726 \\
0,1530 & 1,0000 & 1,0000 & 1,0000 & 1,0000 & 1,0000 & 1,0000 & 1,0000 & 0,9370 & 0,8870 & 0,8570 & 0,8470 \\
0,2435 & 1,0000 & 1,0000 & 1,0000 & 1,0000 & 1,0000 & 1,0000 & 0,9165 & 0,8465 & 0,7965 & 0,7665 & 0,7565 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0,0069 & 1,0000 & 1,0000 & 1,0000 & 1,0000 & 1,0000 & 1,0000 & 1,0000 & 1,0000 & 1,0000 & 1,0000 & 0,9931 \\
0,0045 & 1,0000 & 1,0000 & 1,0000 & 1,0000 & 1,0000 & 1,0000 & 1,0000 & 1,0000 & 1,0000 & 1,0000 & 0,9955
\end{array}\right],$$

$$\ldots \quad \ldots \quad \ldots$$

$$R_8=\left[\begin{array}{c|ccccccccccc}
 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ \hline
0,2252 & 0,7748 & 0,7748 & 0,7748 & 0,7748 & 0,7748 & 0,7748 & 0,7748 & 0,7748 & 0,7748 & 0,7748 & 1,0000 \\
0,1913 & 0,8087 & 0,8087 & 0,8087 & 0,8087 & 0,8087 & 0,8087 & 0,8087 & 0,8087 & 0,8087 & 0,8087 & 1,0000 \\
0,1160 & 0,8840 & 0,8840 & 0,8840 & 0,8840 & 0,8840 & 0,8840 & 0,8840 & 0,8840 & 0,8840 & 0,8840 & 1,0000 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0,8334 & 0,1666 & 0,1666 & 0,1666 & 0,1666 & 0,1666 & 0,1666 & 0,1666 & 0,1666 & 0,1666 & 0,1666 & 1,0000 \\
0,8959 & 0,1041 & 0,1041 & 0,1041 & 0,1041 & 0,1041 & 0,1041 & 0,1041 & 0,1041 & 0,1041 & 0,1041 & 1,0000
\end{array}\right].$$

В результате пересечения $R_1 \cap R_2 \cap \ldots \cap R_8$ получено общее функциональное решение в виде следующей матрицы:

$$R=\left[\begin{array}{c|ccccccccccc}
 & 0 & 0,1 & 0,2 & 0,3 & 0,4 & 0,5 & 0,6 & 0,7 & 0,8 & 0,9 & 1 \\ \hline
x_1=25090,5 & 0,0114 & 0,1114 & 0,2114 & 0,3114 & 0,4114 & 0,5114 & 0,6114 & 0,5610 & 0,4610 & 0,3610 & 0,2610 \\
x_2=24987,5 & 0,0018 & 0,1018 & 0,2018 & 0,3018 & 0,4018 & 0,5018 & 0,6018 & 0,5094 & 0,4094 & 0,3094 & 0,2094 \\
x_3=24700,2 & 0,0193 & 0,1193 & 0,2193 & 0,3193 & 0,4193 & 0,5193 & 0,4878 & 0,3878 & 0,2878 & 0,1878 & 0,0878 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
x_{332}=26346,0 & 0,0002 & 0,0102 & 0,0402 & 0,0902 & 0,1602 & 0,1666 & 0,1666 & 0,1666 & 0,1666 & 0,1666 & 0,8357 \\
x_{333}=26496,7 & 0,0123 & 0,0223 & 0,0523 & 0,1023 & 0,1041 & 0,1041 & 0,1041 & 0,1041 & 0,1041 & 0,1041 & 0,8740
\end{array}\right],$$

которая отражает причинно-следственную связь между признаками локализации ИД $x(t)$, с одной стороны, и, собственно, оценкой величины DJI, с другой. В данном случае нечёткая интерпретация t-ой ИД $A_t\,(t=1\div 333)$ находится посредством правила композиционного вывода: $A_t = G_t \circ R \;\; (t=1\div 333)$, где G_t является отображением t-ой ИД в виде нечёткого подмножества универсума U. Выбирая композиционное правило $\mu_{At}(i)=\max\{\min\{\mu_{Gt}(x),\, \mu_R(i)\}\}$ и полагая $\mu_{Gt}(x)=0$, если $x \neq x_t$ и $\mu_{Gt}(x)=1$, если $x = x_t$, имеем $\mu_{At}(i)=\mu_R(x_t,\, i)$. Это означает, что A_t является нечёткой интерпретацией t-ой ИД на универсуме I, соответствующие значения ФП расположены на t-ой строке матрицы R. В частности, нечётким аналогом ИД $x_1=25090{,}5$ является НМ (1-я строка матрицы R): A_1={0,0114/0; 0,1114/0,1; 0,2114/0,2; 0,3114/0,3; 0,4114/0,4; 0,5114/0,5; 0,6114/0,6; 0,5610/0,7; 0,4610/0,8; 0,3610/0,9; 0,2610/1}.

Таким образом, по числу данных ВР DJI получены нечёткие подмножества универсума I, которые, собственно, и описывают эти ИД (табл. 1). Очевидно, что это слишком большой набор множеств, чтобы можно было на его основе применять известные нечёткие прогностические модели ВР. Поэтому необходимо существенно сузить список НМ.

Таблица 1 – Детализированный нечёткий аналог ВР DJI

Дата	НМ	Значения ФП НМ универсума I					ТО НМ
		0	0,1	…	0,9	1	
15.06.2018	A_1	0,0114	0,1114	…	0,3610	0,2610	0,6062
18.06.2018	A_2	0,0018	0,1018	…	0,3094	0,2094	0,5939
19.06.2018	A_3	0,0193	0,1193	…	0,1878	0,0878	0,5330
20.06.2018	A4	0,0273	0,1273	…	0,1735	0,0735	0,5219
21.06.2018	A5	0,0809	0,1809	…	0,1229	0,0229	0,4723
… … …							
03.10.2019	A328	0,0057	0,0157		0,2358	0,7916	0,9048
04.10.2019	A329	0,0252	0,0352	…	0,0769	0,8907	0,9627
07.10.2019	A330	0,0098	0,0198	…	0,1113	0,8697	0,9513
08.10.2019	A331	0,0098	0,0198	…	0,2546	0,7791	0,8959
09.10.2019	A332	0,0002	0,0102	…	0,1666	0,8357	0,9321
10.10.2019	A333	0,0123	0,0223	…	0,1041	0,8740	0,9534

Приведённые в табл. 1 точечные оценки нечётких множеств (ТО НМ) являются дефаззифицированными значениями нечётких аналогов ИД $A_t(t = 1 \div 333)$, которые в масштабе отрезка [0, 1] условно восстанавливают конфигурацию ВР DJI так, как это показано на рис. 3.

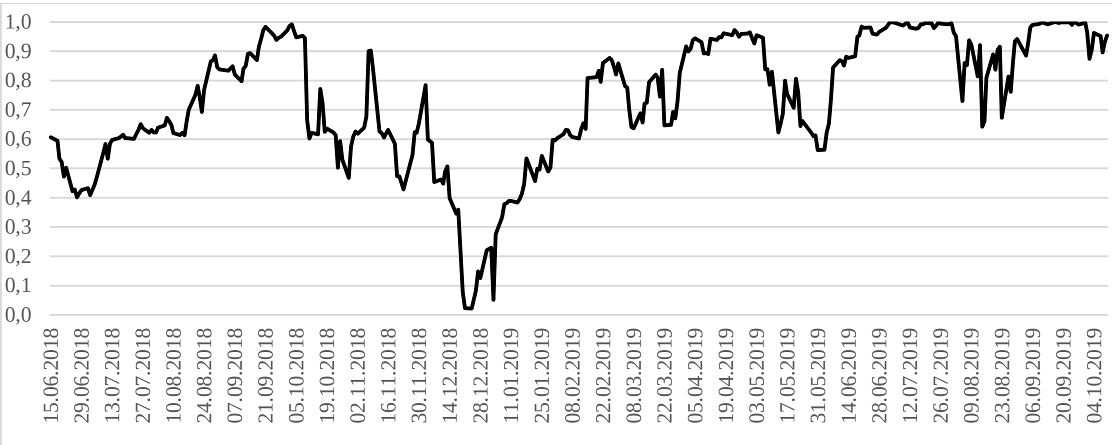

Рисунок 3 – ВР DJI в нотации ТО НМ

ТО НМ устанавливаются согласно следующим рассуждениям. Для нечёткого подмножества универсума, то есть в нашем случае для $A \subset I$ определяются α-уровневые множества ($\alpha \in [0, 1]$): $A_\alpha = \{i | \mu_A(i) \geq \alpha,\ i \in I\}$. Далее, для каждого A_α вычисляется соответствующая мощность $M(A_\alpha)$ – среднее число элементов по формуле $M(A_\alpha) = \sum_{k=1}^{n} u_k / n,\ u_k \in A_\alpha$. В итоге, ТО для НМ A устанавливается по формуле $F(A) = (1/\alpha_{\max}) \int_0^{\alpha_{\max}} M(A_\alpha) d\alpha$, где $\alpha_{\max}$ – максимальное значение ФП НМ A. В частности, для НМ A_1 имеем:

- для $0<\alpha<0{,}0114$, $\Delta\alpha=0{,}0114$, $A_{1\alpha}=\{0; 0{,}1; \ldots; 0{,}8; 0{,}9; 1\}$, $M(A_{1\alpha})=0{,}50$;
- для $0{,}0114<\alpha<0{,}1114$, $\Delta\alpha=0{,}1$, $A_{1\alpha}=\{0{,}1; 0{,}2; \ldots; 0{,}9; 1\}$, $M(A_{1\alpha})=0{,}55$;
- для $0{,}1114<\alpha<0{,}2114$, $\Delta\alpha=0{,}1$, $A_{1\alpha}=\{0{,}2; 0{,}3; \ldots; 0{,}9; 1\}$, $M(A_{1\alpha})=0{,}60$;
- для $0{,}2114<\alpha<0{,}2610$, $\Delta\alpha=0{,}0495$, $A_{1\alpha}=\{0{,}3; 0{,}4;\ldots;0{,}9; 1\}$, $M(A_{1\alpha})=0{,}65$;
- для $0{,}2610<\alpha<0{,}3114$,$\Delta\alpha=0{,}0505$,$A_{1\alpha}=\{0{,}3;0{,}4;\ldots;0{,}8;0{,}9\}$, $M(A_{1\alpha})=0{,}60$;

• для $0,3114<\alpha<0,3610$, $\Delta\alpha=0,0495$, $A_{1\alpha}=\{0,4; 0,5;\ldots; 0,9\}$, $M(A_{1\alpha})=0,65$;
• для $0,3610<\alpha<0,4114$, $\Delta\alpha=0,0505$, $A_{1\alpha}=\{0,4; 0,5; \ldots; 0,8\}$, $M(A_{1\alpha})=0,60$;
• для $0,4114<\alpha<0,4610$, $\Delta\alpha=0,0495$, $A_{\alpha}=\{0,5; 0,6; 0,7; 0,8\}$, $M(A_{1\alpha})=0,65$;
• для $0,4610<\alpha<0,5114$, $\Delta\alpha=0,0505$, $A_{1\alpha}=\{0,5; 0,6; 0,7\}$, $M(A_{1\alpha})=0,60$;
• для $0,5114<\alpha<0,5610$, $\Delta\alpha=0,0495$, $A_{1\alpha}=\{0,6; 0,7\}$, $M(A_{1\alpha})=0,65$;
• для $0,5610<\alpha<0,6114$, $\Delta\alpha=0,0505$, $A_{1\alpha}=\{0,6\}$, $M(A_{1\alpha})=0,60$.

$$F(A_1)=[0,0114\cdot 0,5+0,1\cdot 0,55+\ldots+0,0495\cdot 0,65+0,0505\cdot 0,60]/0,6114=0,6062.$$

Теперь на базе НМ, составленных по числу ИД, сформируем критерии оценки показателя DJI. Пусть число таких критериев будет 10. Разобьём все НМ $A_t (t=1\div 333)$ на 10 групп путём следующих правил: «Если ТО A_t из интервала (0,1(k-1), 0,1k], то A_t из k-ой группы», где $k=1\div 10$. В этом случае k-ым критерием оценки показателя DJI будем считать пересечение всех НМ, попавших в k-ю группу, то есть НМ из k-ой группы, объединенные союзом и/или логической операцией «И». Например, в состав 1-ой группы входят НМ, чьи ТО располагаются в интервале (0, 0,1] (табл. 1). Таковыми являются множества A_{131}, A_{132}, A_{133}, A_{134} и A_{139}. Сведения об этих множествах на универсуме I, включая значения их ФП и ТО, а также НМ $B_1=A_{131}\cap A_{132}\cap A_{133}\cap A_{134}\cap A_{139}$ как 1-го критерия оценки показателя DJI, сведены в табл. 2. Аналогичным образом устанавливаются и остальные обобщённые критерии оценки в виде НМ $B_2\div B_{10}$.

Таблица 2 – Формирование 1-го критерия оценки показателя DJI

НМ	Значения ФП НМ из 1-ой группы на универсуме I						ТО НМ
	0	0,1	0,2	…	0,9	1	
A_{131}	0,8173	0,1960	0,1960	…	0,0106	0,0006	0,0785
A_{132}	0,9138	0,0412	0,0412	…	0,0412	0,0412	0,0226
A_{133}	0,9801	0,0423	0,0423	…	0,0423	0,0423	0,0216
A_{134}	0,8115	0,2051	0,2051	…	0,0113	0,0013	0,0821
A_{139}	0,8644	0,1200	0,1200	…	0,0172	0,0072	0,0514
B_1	0,8115	0,0412	0,0412	…	0,0106	0,0006	0,0210

В итоге, НМ $B_k (k=1\div 10)$, сведения о которых сведены в табл. 3, являются обобщёнными критериями оценки показателя DJI и их условно можно обозначить как B_1 – ЧЕРЕСЧУР НИЗКИЙ, B_2 – ОЧЕНЬ НИЗКИЙ, B_3 – БОЛЕЕ ЧЕМ НИЗКИЙ, B_4 – НИЗКИЙ, B_5 – НИЖЕ СРЕДНЕГО, B_6 – СРЕДНИЙ, B_7 – ВЫШЕ СРЕДНЕГО, B_8 – ВЫСОКИЙ, B_9 – ОЧЕНЬ ВЫСОКИЙ, B_{10} – ЧЕРЕСЧУР ВЫСОКИЙ.

Таблица 3 – Формирование 1-го критерия оценки показателя DJI

НМ	Значения ФП НМ B_k на универсуме I						ТО НМ
	0	0,1	0,2	…	0,9	1	
B_1	0,8115	0,0412	0,0412	…	0,0106	0,0006	0,0210
B_2	0,7193	0,2985	0,2985	…	0,0330	0,0230	0,1308
B_3	0,5875	0,4393	0,4393	…	0,1061	0,0031	0,2376
B_4	0,2361	0,3361	0,4361	…	0,1058	0,0000	0,3830
B_5	0,0466	0,1466	0,2466	…	0,1000	0,0000	0,4767
B_6	0,0000	0,1000	0,2000	…	0,1195	0,0195	0,5098
B_7	0,0000	0,1061	0,2014	…	0,3375	0,2375	0,6237

Продолжение табл. 3

B_8	0,0009	0,0925	0,1225	…	0,4183	0,5679	0,7678
B_9	0,0079	0,0179	0,0479	…	0,2463	0,6597	0,8810
B_{10}	0,0001	0,0010	0,0010	…	0,0010	0,7883	0,9994

Таким образом, если ТО НМ $A_t(t = 1 \div 333)$ попадает в k -ый интервал, то описываемый этим множеством показатель DJI будет оцениваться качественным критерием B_k и отражаться соответствующим НМ. При применении данного подхода получен более общий нечёткий аналог ВР, который сведён в табл. 4. Там же отражен нечёткий ВР с применением критерия $C_i\,(k = 1 \div 7)$.

Таблица 4 – Нечёткие временные ряды DJI

Дата	DJI	Интервал включения	Нечёткий аналог DJI		Нечёткая модель ВР DJI:	
			Обозначение	ТО НМ	с применением $C_i\,(k = 1 \div 7)$	с применением $B_k\,(k = 1 \div 10)$
15.06.2018	25090,5	u_5	A_1	0,6062	C_5	B_7
18.06.2018	24987,5	u_5	A_2	0,5939	C_5	B_6
19.06.2018	24700,2	u_5	A_3	0,5330	C_5	B_6
20.06.2018	24657,8	u_5	A_4	0,5219	C_5	B_6
21.06.2018	24461,7	u_4	A_5	0,4723	C_4	B_5
22.06.2018	24580,9	u_5	A_6	0,5024	C_5	B_6
… … …						
03.10.2019	26201,0	u_7	A_{328}	0,9048	C_7	B_{10}
04.10.2019	26573,7	u_7	A_{329}	0,9627	C_7	B_{10}
07.10.2019	26478,0	u_7	A_{330}	0,9513	C_7	B_{10}
08.10.2019	26164,0	u_7	A_{331}	0,8959	C_7	B_9
09.10.2019	26346,0	u_7	A_{332}	0,9321	C_7	B_{10}
10.10.2019	26496,7	u_7	A_{333}	0,9534	C_7	B_{10}

5. Заключение

Предлагаемый подход к фаззификации ИД апробирован на показателях индустриального индекса Доу-Джонса, которые устанавливаются по результатам ежедневных торгов на фондовой бирже США путём обычного арифметического усреднения составляющих показателей. Последнее позволяет считать ежедневные показания индекса Доу-Джонса как слабоструктурированные, а саму динамику его изменения интерпретировать в виде нечёткого ВР. Поэтому для ИД ВР индекса Доу-Джонса применяется процедура фаззификации, которая в статье реализуется через систему нечёткого вывода, обеспечивающую нахождение значений ФП соответствующих нечётких подмножеств дискретного универсума, покрывающего всю совокупность показателей индекса за более чем годичный период. В статье выбраны всего десять критериев оценки ИД. Однако для имплементации известных моделей прогнозирования в данном конкретном случае число таких критериев недостаточно. В рамках предлагаемого подхода можно легко и быстро увеличить число этих критериев оценки для более детального описания ВР.

Данная работа выполнена при финансовой поддержке Фонда Развития Науки при Президенте Азербайджанской Республики – Грант №EİF/MQM/Elm-Tehsil-1-2016-1(26)-71/15/5).

СПИСОК ИСТОЧНИКОВ

1. Zadeh L.A. The concept of a linguistic variable and its application to approximate reasoning. *Information sciences*. 1975. Vol. 8, N 3. P. 199–249.
2. Song Q., Chissom B.S. Fuzzy time series and its models. *Fuzzy Sets and Systems*. 1993. N 54. P. 269–277.
3. Kumar N., Ahuja S., Kumar V. Fuzzy time series forecasting of wheat production. *International journal on computer science and engineering*. 2010. Vol. 2, N 3. P. 635–640.
4. Chen S.M. Forecasting enrollments based on fuzzy time series. *Fuzzy sets and systems*. 1996. N 81. P. 311–319.
5. Cheng C.H., Chang J.R., Yen C.A. Entropy-based and trapezoid fuzzification fuzzy time series approaches for forecasting IT project cost. *Technological forecasting & social change.* 2006. N 73. P. 524–542.
6. Poulsen J.R. Fuzzy time series forecasting – developing a new forecasting model based on high order fuzzy time series. AAUE: CIS 4, 2009. 67 p.
7. Рзаев Р., Шихалиева Г., Агамалыев М. Моделирование временных рядов на основе нечёткого анализа данных. *Нечеткие системы и мягкие вычисления*. 2014. Т. 9. № 1. С. 39–86.
8. Промышленный индекс Доу-Джонса. URL: https://ru.tradingview. com/symbols/DJ-DJI/ (дата обращения: 11.03.2020).

Стаття надійшла до редакції 04.05.2020