

Новая гауссовская система нечёткого логического вывода типа Такаги-Сугено-Канга с весом главного компонента

Нгуен Тхи Тху Зунг

Аннотация—В настоящее время нечеткие системы вывода оказывают чрезвычайно эффективную поддержку в решении многих задач в практической деятельности. Среди таких систем стоит отметить мощные характеристики и применение нечеткой системы Такаги-Сугено-Канга (ТСК). В данном исследовании предложена новая гауссовская нечеткая система ТСКа с использованием метода анализа главных компонент для уменьшения емкости системы нечетких логических правил, когда количество входных показателей относительно много, и в то же время модель имеет улучшения при рассмотрении весовых значений входных показателей путем определения пропорции вклада входной информации. Модель использует подход минимизации энтропии (МЕРА) для обеспечения эффективной поддержки процесса фазификации ряда входных данных. Предложенная модель применяется для прогнозирования индекса социально-экономического развития 63 провинций Вьетнама, в качестве входных данных используются показатели социально-экономического развития 63 провинций в 2019 году. Результаты прогнозирования предложенной системы оцениваются и анализируются путем сравнения прогнозных и фактических значений, а также оценки значений MSE, RMSE, MAPE и CORR между моделью, использующей весовые коэффициенты, и моделью, не использующей весовые коэффициенты. Предложенная модель с использованием весовых коэффициентов имеет более высокую производительность.

Ключевые слова— Нечеткая система вывода Такаги-Сугено-Канга, принцип минимизации энтропии (МЕРА), весовые коэффициенты главных компонент, сокращенные критерии, нечеткая система на основе правил, социально-экономический индекс.

I. ВВЕДЕНИЕ

В настоящее время все большинство физических динамических систем в реальности являются нелинейными и сложными. Развитие нечеткого моделирования является большим прорывом, помогающим решать многие сложные задачи, основанные на разработке или математическом моделировании, подходящим и эффективным способом. Нечеткая система вывода является одним из выдающихся достижений в нечетком моделировании, основанном на описании нечетких правил ЕСЛИ-ТО,

которые представляют локальные отношения ввода-вывода нелинейной системы. Нечеткая система вывода, разработанная Заде, зародилась и привлекла внимание большого числа исследователей в последние три десятилетия [1]. Нечеткие системы вывода получают довольно сильное развитие [2]–[3], находя разнообразные применения в таких задачах, как прогнозирование временных рядов, поддержка принятия решений, классификация, управление и моделирование [4], [5], в различных областях, таких как: машиностроение [6], химия [7], биология [8], финансы [9], технологии [10]–[13]. В целом, нечеткая система вывода включает в себя: систему нечетких правил, связанную с нечеткими множествами, нечеткую логику, нечеткий вывод. Системы нечетких правил играют важную роль в моделировании и рассуждениях системы нечетких выводах. В настоящее время разработаны различные типы систем на основе нечетких правил, такие как: генетические нечеткие системы (ГНС), нейро-нечеткие системы (ННС), иерархические нечеткие системы (ИНС), эволюционирующие нечеткие системы (ЭНС) и многоцелевые нечеткие системы (МНС), однако все они развиваются из основных типов систем нечетких правил, таких как: Нечеткие правила Мамдани [14], нечеткие правила Такаги-Сугено-Канга (ТСКа) [15], система правил Сукамото и система правил Синглтона. Среди них нечеткая система вывода ТСКа более популярна, поскольку модель ТСКа очень полезна для моделирования систем, которые трудно описать с помощью чисто математических моделей. Система ТСКа оказывается наиболее надежной и имеет наибольшее количество применений [16]–[18].

Нечеткая система ТСКа исследована и имеет множество улучшений для повышения эффективности прогнозирования. В работе [19] метод нечеткой кластеризации С-средних используется для определения количества нечетких правил и параметров функции принадлежности. Далее, в работах [20], [21] преимущества нечеткой кластеризации С-средних используются для разработки нечеткого контроллера на основе обучающих данных. Кроме того, в работе [22] предложен новый подход к улучшению локальной и глобальной аппроксимации системы ТСКа, однако при этом возросла вычислительная сложность. Кроме того, по результатам работы [23] построен, спроектирован и использован для навигации мобильных роботов нечеткий логический контроллер (НЛК) на основе нейро-нечеткой системы Такаги-Сугено с

Статья получена 20 декабря 2023.

Нгуен Тхи Тху Зунг, аспирантка, Россия, Санкт-Петербург, Санкт-Петербургский Политехнический университет Петра Великого (e-mail: thudung.mta.tb@gmail.com; ORCID: <https://orcid.org/0000-0001-9206-5968>).

использованием дифференциальных функций принадлежности. В работе [24] рассматривается решение задачи интервальной аппроксимации Л-нормы с использованием метода ЛП и метода нечеткой логики Такаги-Сугено. Наряду с этим в работе [25] представлен алгоритм, использующий нечеткую систему Такаги-Сугено с инкрементальным обучением на каждом временном шаге и сочетающий методы корреляции и частичной автокорреляции между переменными для определения количества входной переменной. В работе [26] рассматривается проблема корректировки весов правил в нейронной системе нечеткой логики ТСКА и предлагаются некоторые эвристические методы для расчета весов правил. Кроме того, в работе [27] предложен метод оптимизации производительности нечеткой нейронной системы ТСКА с помощью Мини-пакетный градиентный спуск, модифицированный AdaBound (MBGD-A) для наблюдения за точностью классификации, выполняемой ТСКА с помощью MBGD-A. Системам нечеткого вывода ТСКА в сочетании с нейросетевыми моделями уделяется достаточно много внимания, например, модели адаптивной нейро-нечеткой системы вывода (ANFIS) [28]–[34].

Однако, как оказалось, недостатком систем нечеткого вывода является то, что количество нечетких правил достаточно велико. В предыдущих исследованиях системы интерполяции нечеткого вывода обеспечивали очень эффективную поддержку [35], но при большом количестве входных критериев это потребовало бы большой работы [36]. Кроме того, насколько известно автору, предыдущие исследования в основном посвящены оценке весов нечетких правил в системах нечеткой вывода, но очень мало исследований рассматривают веса входных показателей [37]. Поэтому в данном исследовании модель предлагает уменьшить размерность входных данных с помощью анализа главных компонент, чтобы значительно сократить количество мерностей без потери информации. После этого веса показателей определяются на основе доли их вклада в исходные данные. Кроме того, в модели используется принцип минимизации энтропии для эффективной поддержки фазификации. Предложенная модель применяется для прогнозирования индекса социально-экономического развития 63 провинций Вьетнама в 2019 году. Исходные данные собраны из 120 показателей социально-экономического развития провинций Вьетнама по данным Главного статистического управления Вьетнама.

Данная работа состоит из 5 основных разделов. Сначала дается введение в историю исследования. В следующем разделе представлены математические теоретические основы системы нечеткого вывода. В разделе 3 описана предлагаемая модель и алгоритм ее реализации. В разделе 4 предложенная модель применяется к реальным задачам и анализируется эффективность предложенной модели. В последней части приводятся заключения.

II. ТЕОРЕТИЧЕСКИЕ ОСНОВЫ

A. Нечеткая гауссовская система вывода по типу Такаги-Сугено-Канга

Модель системы нечетких выводов Такаги-Сугено-Канга (ТСКА) — это модель, основанная на нечетких правилах, которая подходит для аппроксимации многих систем и функций [15]. В нечеткой модели ТСКА следствия правил обычно принимаются либо за четкие числа, либо за линейные функции входов. Математическая основа модели ТСКА в общем виде представляется следующим образом.

Пусть $x \in X$, $X = \{x_1, x_2, \dots, x_i, \dots, x_m\}$ будут входные переменные, определенные во вселенной $X_1, X_2, \dots, X_i, \dots, X_m$, и пусть y будет выходная переменная, определенная во вселенной Y .

Тогда k -го правило типа Такаги-Сугено-Канга определяется следующим образом:

ЕСЛИ x_1 является $A_{1,k}$ И x_2 является $A_{2,k}$ И ... И x_i является $A_{i,k}$ И ... И x_m является $A_{m,k}$ ТО $y_k^m = a_{1,k}^m x_1 + a_{2,k}^m x_2 + \dots + a_{i,k}^m x_i + \dots + a_{n,k}^m x_n + b^m$, $i = \overline{1, n}$.

В данной работе учитываются веса каждой исходной мерности. Пусть w_i , $i = \overline{1, m}$ — вес мерности, соответствующей m размерности входной величины x_i , $i = \overline{1, m}$. Тогда вес выходного уровня y_k каждого k -го правила "ЕСЛИ - ТО" R_k равен

$$\sigma_k = \min(\mu_{A_{i,k}} \times w_i) \quad (1)$$

Выход y^m всех N правил ЕСЛИ-ТО, вычисляется следующим образом:

$$y = \frac{1}{N} \sum_{k=1}^N (y_k \cdot \sigma_k) \quad (2)$$

B. Предложенный метод

В данной работе предлагается новая модель, которая учитывает веса главных компонент каждого входного критерия, определяя пропорцию вклада главных компонент в исходные данные, и в то же время нечеткие правила минимизируются путем уменьшения исходной размерности на основе метода анализа главных компонент. Кроме того, модель использует принцип минимизации энтропии для процесса фазификации, разделяя последовательности входных данных на соответствующие нечеткие множества в соответствии с наименьшим значением частичной энтропии. Схема предлагаемой модели показана на рисунке 1.

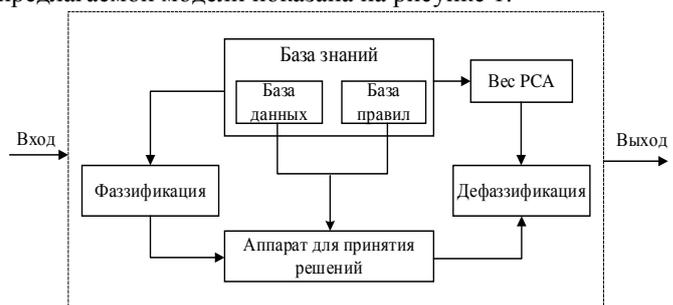


Рис. 1. Предложенная нечеткая система управления.

а. Метод главных компонент

Суть метода главных компонент (PCA) заключается в представлении p точек данных в виде облака в k dimensional space мерном пространстве с центром тяжести, расположенным в начале координат. PCA пытается найти подпространство k мерности α в \mathbb{Y}^n таком, что $P_{\alpha} X$ - ортогональные проекции p точек на α имеют максимальную дисперсию [38], [39].

Пусть есть n наблюдения, соответствующие n точкам в k - мерном пространстве, тогда существует k -мерный вектор в виде $\bar{X} = (X_1, X_2, \dots, X_k)^T$. Входные значения n наблюдений имеют вид $X = [x_{ij}]^{n \times k}$, $i = \overline{1, n}$, $j = \overline{1, k}$.

Тогда нормированный вектор имеет вид

$$\bar{X}_j = X - M X = [\bar{x}_j]^{n \times k},$$

где $\bar{x}_j = \frac{x_{ij} - \bar{x}_j}{\sqrt{\sigma_j}}$, и значение стандартного

отклонения j -го вектора $\sigma_j = \frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2$.

Ковариационная матрица имеет следующий вид:

$$C = M(\bar{X} \bar{X}^T) = [c_{jj}]^{k \times k} \tag{3}$$

где $c_{jj} = \frac{1}{n} \sum_{i=1}^n (\bar{x}_j - \bar{\bar{x}}_j)(\bar{x}_j - \bar{\bar{x}}_j)$, $j, j' = \overline{1, k}$, $i = \overline{1, n}$.

Предположим, что собственные значения ковариационной матрицы равны $l_1 > \dots > l_l > \dots > l_k$. Пропорция собственных значений вычисляется по

$\theta = [q_j]^{k \times 1}$, с $q_l = l_l \frac{1}{\sum_{l=1}^k l_l}$. Тогда соответствующие

собственные векторы, определенные на основе уравнения $(C - l_l E)v_l = 0$, имеют вид $V = (v_1, \dots, v_l, \dots, v_k)$, $l = \overline{1, k}$.

Тогда вес критерия определяется по значению собственных векторов следующим образом:

$$W = V\theta = [w_l], w_l = |v_l| \cdot q_l \tag{4}$$

Количество сохраняемых главных компонент рассчитывается в порядке убывания значения l , удовлетворяющего определенному уровню входной информации. Тогда

$$J_{k\bar{y}} = \frac{\sum_{j=1}^{k\bar{y}} l_j}{\sum_{j=1}^k l_j} \tag{5}$$

- пропорция дисперсии, вносимой первыми $k\bar{y}$ главными компонентами.

Метод главных компонент сохраняет изменчивость исходных переменных настолько, насколько это возможно по значению $J_{k\bar{y}}$.

б. Принципиальный подход с минимизацией энтропии (МЕРА)

Основная идея метода заключается в разделении рядов данных выборки по точкам отсечения, определяемым путем расчета точек, имеющих наименьшую энтропию [40]. Предположим, что значения определяемой строки находятся в диапазоне от x_1 до x_2 . Уравнение энтропии с каждым значением x

записывается для регионов $[x_1, x_1 + x]$ и $[x_1 + x, x_2]$, обозначает первую регион как p , вторую - q . Энтропия с каждым значением x в области между x_1 и x_2 выражается следующим образом [37]:

$$S(x) = p(x)S_p(x) + q(x)S_q(x)$$

где

$$S_p(x) = -[p_1(x)\ln p_1(x) + p_2(x)\ln p_2(x)];$$

$$S_q(x) = -[q_1(x)\ln q_1(x) + q_2(x)\ln q_2(x)]$$

где $p_k(x)$ и $q_k(x)$ - условные вероятности того, что выборка класса k находится в области соответственно $[x_1, x_1 + x]$ и $[x_1 + x, x_2]$; при $p(x) + q(x) = 1$ где $p(x)$ и $q(x)$ - вероятность того, что все образцы присутствуют в области соответственно $[x_1, x_1 + x]$ и $[x_1 + x, x_2]$.

Учитывая значения x в регионе $[x_1, x_2]$, вычисляются значения энтропии для каждой точки x . Значение x содержащее минимальную энтропию, будет выбрано в качестве главного отрезка (PRI) для разделения региона $[x_1, x_2]$ на два, соответствующих двум областям двух нечетких множеств. Процесс деления будет продолжаться с каждым меньшим регионом в зависимости от количества выбранных нечетких групп.

III. ЭКСПЕРИМЕНТЫ И АНАЛИЗ

Предложенная модель экспериментально используется для прогнозирования индекса S - социально-экономического развития 63 провинций Вьетнама, используя данные социально-экономического развития по критериям в качестве исходных данных. Социально-экономическое положение 63 провинций по 20 основным критериям, разделенным на 120 малых социально-экономических критериев в 2019 году. В соответствии с шагами, указанными в вышеприведенном разделе, процесс обработки осуществляется следующим образом

A. Эксперименты

Шаг 1. Входные данные

По данным Главного статистического управления Вьетнама, критерии оценки социально-экономического развития регионов включают 20 основных критериев, разделенных на 120 подкритериев. Данные по первому из 20 критериев для некоторых из 63 провинций представлены в таблице 1.

ТАБЛИЦА 1. КРИТЕРИЙ 1 "ТЕРРИТОРИЯ И НАСЕЛЕНИЕ" (ФРАГМЕНТ)

	Q1.1	Q1.2	Q1.3	Q1.4	Q1.5	Q1.6	Q1.7
HN	335,9	2410	98,32	19,1	2,24	5,7	2,27
VP	123,6	934	99,32	17,2	2,39	5,9	1,45
BN	82,3	1676	97,62	24,8	2,53	5,6	3,08
QN	617,8	214	103,48	16,2	2,24	5,5	1,61
HD	166,8	1137	98,8	17,4	2,48	6,6	1,02
HP	156,2	1302	98,74	15,9	2,2	7,3	0,83
HY	93	1350	100,25	17,1	2,4	6,3	1,08
TB	158,6	1174	94,84	15,2	2,43	7,5	0,43

Шаг 2. Сокращение числа критериев.

Ковариационная матрица имеет собственные значения, показанные на рисунке 2. В данной работе

выбран уровень сохранения информации 95 %, что соответствует сохранению 30 главных компонент из общего 120 компонент. Пояснение уровня информативности главных компонент представлено на рисунке 3.

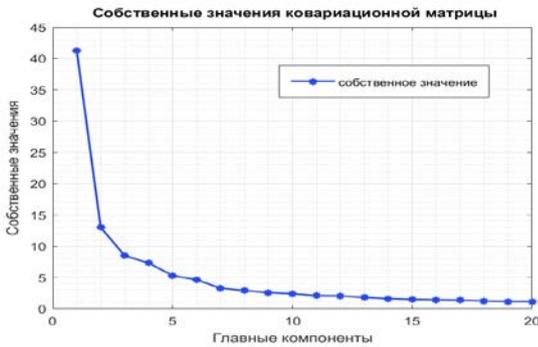


Рис. 2. Первые 20 собственных значений ковариационной матрицы.

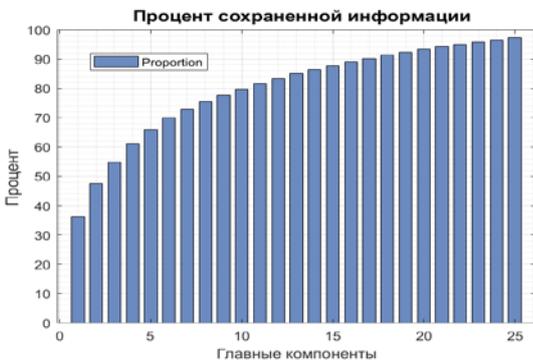


Рис. 3. Пропорция интерпретации информации для 30 сохранившихся главных компонент

Далее определяются наблюдения 63 провинций по 30 новым критериям, результаты представлены в таблице 2.

ТАБЛИЦА 2. НАБЛЮДАЕМЫЕ ЗНАЧЕНИЯ ДЛЯ 30 НОВЫХ ГЛАВНЫХ КОМПОНЕНТ (ФРАГМЕНТ)

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
HN	-30,7	5,5	14,3	-1,4	-5,2	2,3	0,0
VP	-0,2	-2,6	0,6	3,5	1,9	1,8	-3,5
BN	-2,3	-2,3	0,9	6,4	0,3	1,6	-0,9
QN	-1,3	0,0	-0,5	2,2	1,9	1,4	0,1
HD	-2,4	-2,7	2,2	1,6	0,7	2,3	-1,6
HP	-4,7	-3,2	1,8	2,2	0,8	2,5	-0,6
HY	-0,3	-2,9	0,5	2,2	1,4	2,2	-2,2
TV	-0,8	-3,5	1,1	-0,2	2,0	2,3	-3,3

Шаг 3. Определение веса критериев

Таким образом, после сокращения числа критериев получается 30 вместо 120, при этом потеря информации незначительна, всего около 5 %. Для определения весов 30 новых критериев в данной статье проводится повторное PCA. Результаты полученных значений собственных векторов 30 главных компонент вместе с долей их интерпретации информации приведены в таблице 3

ТАБЛИЦА 3. СОБСТВЕННЫЕ ВЕКТОРЫ И ДОЛИ ИНТЕРПРЕТАЦИИ ИНФОРМАЦИИ ИЗ 30 ГЛАВНЫХ КОМПОНЕНТ (ФРАГМЕНТ)

	DPC1	DPC2	DPC3	DPC4	DPC5	...	Весы (w _j)
Доли.	0,361	0,114	0,074	0,064	0,046	...	-
PC1	-0,119	-0,235	0,048	0,198	-0,146	...	0,138
PC2	-0,199	-0,175	-0,037	0,117	0,198	...	0,158
PC3	0,158	0,159	-0,032	-0,419	-0,350	...	0,167
PC4	-0,092	0,139	0,022	0,067	-0,086	...	0,116
PC5	0,145	-0,026	-0,030	-0,100	-0,075	...	0,127
PC6	-0,033	-0,150	-0,198	0,145	0,057	...	0,101
PC7	-0,246	-0,051	0,022	0,114	0,051	...	0,163

В то же время соответствующие веса для 30 критериев определены по формуле (4). Например, для определения веса первой главной компоненты PC1 имеем:

$$w_1 = 0.119 * 0.361 + 0.235 * 0.114 + 0.048 * 0.074 + 0.198 * 0.064 + \dots = 0.138$$

Аналогичным образом определены значения весов для других 29 главных компонент, результаты приведены в таблице 3. После проведения PCA во второй раз обновлены наблюдаемые значения для 30 главных компонент, полученные результаты приведены в таблице 4.

ТАБЛИЦА 4. НАБЛЮДАЕМЫЕ ЗНАЧЕНИЯ ДЛЯ 30 НОВЫХ ГЛАВНЫХ КОМПОНЕНТ ПОСЛЕ ВТОРОГО РАЗА PCA (ФРАГМЕНТ)

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
HN	1,31	1,35	-0,81	-2,33	-0,32	-0,31	0,33
VP	0,48	0,42	-0,24	0,62	0,64	0,61	0,13
BN	-0,17	2,28	-2,43	0,60	-0,15	1,62	0,40
QN	-0,51	0,23	1,23	0,08	-1,48	0,43	-0,10
HD	-0,16	-0,48	0,23	0,05	-2,07	0,09	2,03
HP	-0,06	-2,33	0,30	-1,77	0,23	0,93	-0,25
HY	-0,05	0,02	0,23	-1,12	0,45	0,98	-0,24
TV	0,57	-0,39	0,20	-0,97	-0,86	0,43	0,86

Шаг 4. Фаззификация

В данной работе используется подход MEPA для построения нечетких множеств для исходных данных 30 критериев. В частности, для критерия 1 имеем данные находящиеся между $min(PC1) = -2.52$ и $max(PC1) = 3.07$

. Основываясь на алгоритме MEPA, мы находим точки отсечения, с $PRI = 0.42$, $SEC1 = -0.69$, $SEC2 = 2.46$. В данной работе модель использует гауссовскую нечеткую функцию принадлежности для моделирования нечетких множеств, фаззифицируя ряд исходных данных по каждому критерию в 3 нечетких множества: низкое (Н), среднее (С), высокое (В). Таким способом функция принадлежности первого главного компонента (PC1) определяется следующим образом и представлена на рисунке 2:

$$m_{\text{низкое}}^{PC1}(x) = \exp\left\{-\frac{(x + 0.69)^2}{2c^2}\right\}$$

$$m_{\text{среднее}}^{PC1}(x) = \exp\left\{-\frac{(x - 0.42)^2}{2c^2}\right\}$$

$$m_{\text{высокое}}^{PC1}(x) = \exp\left\{-\frac{(x - 2.46)^2}{2c^2}\right\}$$



Рис. 4. Гауссовская функция принадлежности для PC1

Аналогично, оставшиеся главные компоненты также делятся на 3 нечеткие гауссовские функции принадлежности: "Низкий", "Средний", "Высокий" с различными точками отсечения, основанными на использовании метода МЕРА.

Шаг 5. Построение системы правил

В дальнейшем, используя значения 2019 для обучения, строится система нечетких правил. В соответствии по принципу Менделя для построения нечетких правил, полученная система правил включает 63 правила [41]. Первое правило имеет следующий вид

R1: ЕСЛИ PC1 является «С» И PC2 является «С» И

PC3 является «С» И PC4 является «Н» И PC5 является «В» И PC6 является «В» И PC7 является «С» И PC8 является «С» И PC9 является «В» И PC10 является L И PC11 является «Н» И PC12 является «В» И PC13 является «Н» И PC14 является «С» И PC15 является «Н» И PC16 является «С» И PC17 является «Н» И PC18 является «Н» И PC19 является «Н» И PC20 является «С» И PC21 является «Н» И PC22 является «С» И PC23 является «В» И PC24 является «С» И PC25 является «В» И PC26 является «С» И PC27 является «Н» И PC28 является «Н» И PC29 является «С» И PC30 является «В» THEN S19 является y_1 .

Шаг 6. Дефазификация

На основе системы из 63 правил ЕСЛИ-ТО, определенных выше, для каждого векторного входного значения определяется выходное прогнозируемое значение по формуле (2). Например, для входного значения $x = (1.31, 1.35, - 0.816, - 2.33, - 0.321, \dots)$, Весовые значения нечетких правил определяются оператором \min , и вычисленное выходное значение прогноза равно $y = 78.28$, Процесс прогнозирования показан на рисунке 5.

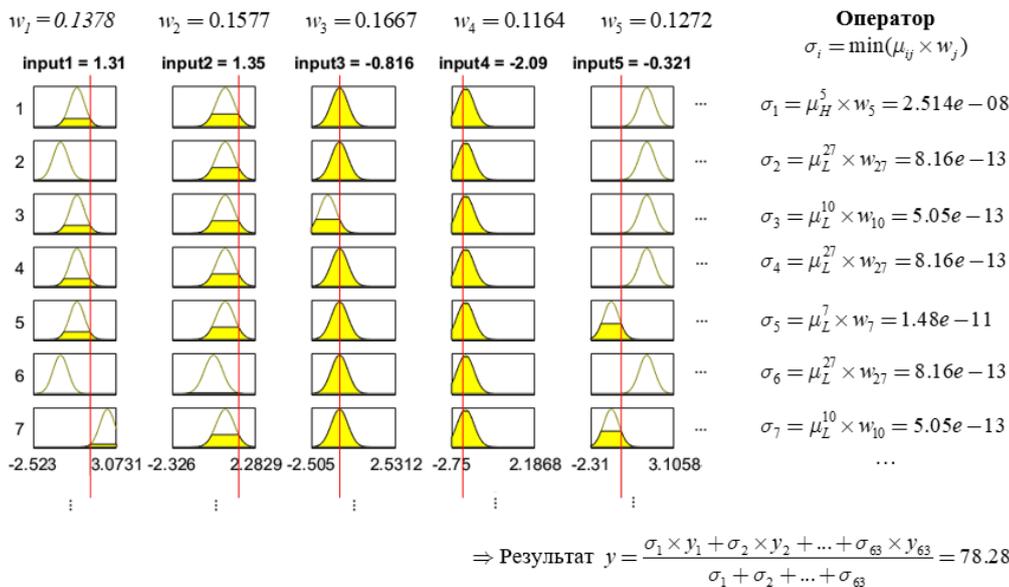


Рис. 5. Процесс дефазификации с входным вектором x

В. Анализ результатов прогнозирования

Результаты прогнозирования по предложенной модели оцениваются путем сравнения фактических и прогнозных значений индекса S, как показано на рисунке 6. Видно, что фактические и прогнозные значения по предложенной модели близки друг к другу. В то же время модель, использующая весовые значения, дает лучшие результаты прогнозирования, чем модель прогнозирования без веса, предсказанные значения по предложенной модели гораздо ближе к реальности.

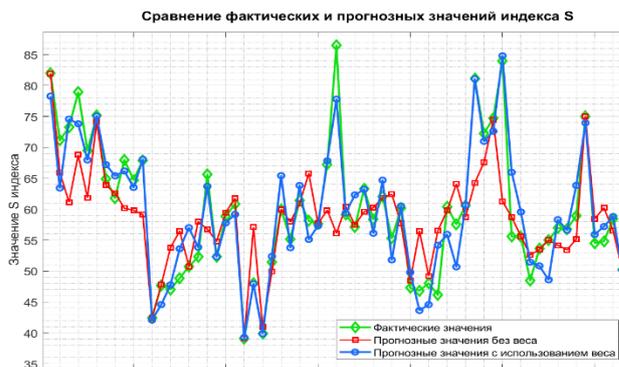


Рис. 6. Сравнение фактических и прогнозируемых значений.

Кроме того, эффективность прогнозирования

оценивается по индексам MSE, RMSE, MAPE и CORR. Для модели без весов значения MSE = 49,35, RMSE = 7,02, MAPE = 6,92 и CORR = 0,77 показывают, что средняя квадратичная ошибка между фактическим и прогнозируемым значением составляет 49,35, среднеквадратичная ошибка составляет 7,02, средняя абсолютная процентная ошибка составляет 6,92, а значение коэффициента корреляции между фактическим и прогнозируемым значениями составляет 0,77.

В предложенной модели с использованием весового значения, значения MSE = 12,1715, RMSE = 3,4888, MAPE = 4,3106 и CORR = 0,945 показывают, что средняя квадратичная ошибка между фактическим и прогнозируемым значением составляет 12,1715, среднеквадратичная ошибка составляет 3,4888, средняя абсолютная процентная ошибка составляет 4,3106, а значение коэффициента корреляции между фактическим и прогнозируемым значением составляет 0,945. Значения ошибок относительно ниже, что доказывает, что предложенная модель с использованием весовых коэффициентов имеет лучшую эффективность прогнозирования, чем модель без использования весовых коэффициентов.

IV. ЗАКЛЮЧЕНИЕ

В данной работе предложен новый метод для гауссовской нечеткой системы вывода Такаги-Сугено-Канга, модель преодолевает проблему традиционной модели вывода Такаги-Сугено-Канга в вычислении весовых значений входных критериев. Используя метод взвешивания главных компонент, определение доли вклада в исходный объем информации тесно связано с их весами. В то же время метод решает проблему о сокращении количества нечетких правил при большой размерности входных данных путем снижения мерности с помощью метода анализа главных компонент, тем самым значительно сокращая количество критериев, со 120 критериев до 30 критериев с потерей всего около 5% информации. Кроме того, предложенный метод сочетает в себе фазсификацию рядов входных данных по каждому критерию с использованием принципа минимизации энтропии (MEPA). Метод протестирован на примере прогнозирования индекса социально-экономического развития S в 2019 году для 63 провинций Вьетнама. Результаты прогнозирования оценены и показывают, что предложенная модель дает хорошие результаты прогнозирования, фактические и прогнозируемые значения относительно близки друг к другу. Кроме того, эффективность модели оценивается по метрикам MSE, RMSE, MAPE и CORR. Результаты анализа показывают, что корреляция между фактическими и предсказанными значениями, выведенными моделью, составляет более 94,5%. Анализ результатов показывает, что предложенная модель с использованием весов для критериев имеет более высокие результаты, чем модель без использования весов.

БЛАГОДАРНОСТИ

Выражаем благодарность учёным и исследователям в рассматриваемой области, предложившим эффективное направление прогнозирования временных рядов, которое является основой для разработки новых улучшений, предложенных в этой работе.

БИБЛИОГРАФИЯ

- [1] L. A. Zadeh, "Fuzzy Sets," *Information and control*, vol. 8, pp. 338–353, 1965.
- [2] O. Castillo and P. Melin, "A review on interval type-2 fuzzy logic applications in intelligent control," *Inf Sci (N Y)*, vol. 279, pp. 615–631, 2014.
- [3] S. H. Liao, "Expert system methodologies and applications - a decade review from 1995 to 2004," *Expert Syst Appl*, vol. 28, no. 1, pp. 93–103, 2005.
- [4] C. C. Lee, "Fuzzy logic in control systems: fuzzy logic controller," *IEEE Trans Syst Man Cybern*, vol. 20, no. 2, pp. 404–418, 1990.
- [5] H. O. Wang, K. Tanaka, and M. F. Griffin, "An approach to fuzzy control of nonlinear systems: Stability and design issues," *IEEE Trans Syst Man Cybern*, vol. 4, no. 1, pp. 14–23, 1996.
- [6] R.-E. Precup and H. Hans, "A survey on industrial applications of fuzzy control," *Comput Ind*, vol. 62, no. 3, pp. 213–226.
- [7] M. Komiyama, K. Yoshimoto, M. Sisido, and K. Ariga, "Chemistry can make strict and fuzzy controls for bio-systems: DNA nanoarchitectonics and cell-macromolecular nanoarchitectonics," *The Chemical Society of Japan*, vol. 90, no. 9, pp. 967–1004, 2017.
- [8] Y. Jin and L. Wang, *Fuzzy Systems in Bioinformatics and Computational biology*, Studies in. Springer, 2009.
- [9] G. Bojadziev, *Fuzzy logic for business, finance, and management*, vol. 23. 2007.
- [10] J. Gomez and D. Dasgupta, "Evolving fuzzy classifiers for intrusion detection," in *2002 IEEE Workshop on inf. Assur*, I. C. Press, Ed., New York, 2002, pp. 321–323.
- [11] S. Elhag, A. Fernandez, A. Bawakid, S. Alshomrani, and F. Herrera, "On the combination of genetic fuzzy systems and pairwise learning for improving detection rates on intrusion detection systems," *Expert Syst Appl*, vol. 42, no. 1, pp. 193–202, 2015.
- [12] A. Lemos, W. Caminhas, and G. Fernando, "Adaptive fault detection and diagnosis using an evolving fuzzy classifier," *Inf Sci (N Y)*, vol. 220, pp. 64–85, 2013.
- [13] P. Melin, O. Mendoza, and O. Castillo, "Face recognition with an improved interval type-2 fuzzy logic sugeno integral and modular neural networks," *IEEE Transactions on Systems, Man and Cybernetics, Part A. Syst. Humans*, vol. 41, no. 5, pp. 1001–1012, 2011.
- [14] E. H. Mamdani, "Application of fuzzy algorithms for control of simple dynamic plant," *Proceedings of the Institution of Electrical Engineers*, vol. 121, no. 12, pp. 1585–1588, 1974.
- [15] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control," *IEEE Trans Syst Man Cybern*, vol. 15, no. 1, pp. 116–132, 1985.
- [16] B. Mohammed and A. Mostafa, "Comparison of Mamdani-type and Sugeno-Type Fuzzy Inference

- Systems for Fuzzy Real Time Scheduling,” *International Journal of Applied Engineering Research*, vol. 11, no. 22, pp. 11071–11075, 2016.
- [17] S. Tina *et al.*, “A comparative Analysis of the Mamdani and Sugeno Fuzzy Inference systems for MPPT of an Islanded PV System,” *Int J Energy Res*, vol. 2023, p. 14, 2023.
- [18] T. Tettey and T. Marwala, “Modelling Conflict: Knowledge Extraction using Bayesian Neural Network and Neuro-fuzzy Models,” *Comput Intell*, no. Mid, 2016.
- [19] S. G. Cao, N. W. Rees, and G. Feng, “Analysis and design for a class of complex control systems - Part I: fuzzy modeling and identification, Automatica,” *Automatica*, vol. 33, no. 6, pp. 1017–1028.
- [20] T. P. Hong and C. Y. Lee, “Induction of fuzzy rules and membership functions from training examples,” *Fuzzy Sets Syst*, vol. 84, no. 1, pp. 33–47, 1996, doi: 10.1016/0165-0114(95)00305-3.
- [21] F. Klawonn and R. Kruse, “Constructing a fuzzy controller from data,” 1997.
- [22] T. A. Johansen, R. Shorten, and R. Murray-Smith, “On the interpretation and identification of dynamic Takagi-Sugeno fuzzy models,” *IEEE Transactions on Fuzzy Systems*, vol. 8, no. 3, pp. 297–313, 2000, doi: 10.1109/91.855918.
- [23] S. Petrovic-Lazarevic, “Neuro-Fuzzy Support of Knowledge Management in Social Regulation,” no. February, pp. 387–400, 2003, doi: 10.1063/1.1503710.
- [24] I. Škrjanc, S. Blažič, and O. Agamennoni, “Interval fuzzy model identification using l_∞ -norm,” *IEEE Transactions on Fuzzy Systems*, vol. 13, no. 5, pp. 561–568, 2005, doi: 10.1109/TFUZZ.2005.856567.
- [25] M. Aqil, I. Kita, A. Yano, and S. Nishiyama, “A Takagi-Sugeno fuzzy system for the prediction of river stage dynamics,” *Jpn Agric Res Q*, vol. 40, no. 4, pp. 369–378, 2006, doi: 10.6090/jarq.40.369.
- [26] K. Simiński, “Rule weights in a neuro-fuzzy system with a hierarchical domain partition,” *International Journal of Applied Mathematics and Computer Science*, vol. 20, no. 2, pp. 337–347, 2010, doi: 10.2478/v10006-010-0025-3.
- [27] S. Surono, G. K. Wen, C. W. Onn, Y. Bin Dasril, A. Y. Astuti, and N. Periasamy, “Effectiveness of Enhanced Takagi Sugeno Kang’S Fuzzy Inference Model,” *J Theor Appl Inf Technol*, vol. 100, no. 8, pp. 2383–2392, 2022.
- [28] C. T. Lin and G. C. S. Lee, “Neural-Network-Based Fuzzy Logic Control and Decision System,” *IEEE Transactions on Computers*, vol. 40, no. 12, pp. 1320–1336, 1991, doi: 10.1109/12.106218.
- [29] Нгуен Тхи Тху Зунг and Л. В. Черненькая, “Дискретизация в моделях прогнозирования нечетких временных рядов,” *Журнал Известия Тульского государственного университета – Технические науки (ТулГУ, г. Тула)*, vol. 8, no. Системный анализ, Управление и обработка информации, pp. 296–304, 2023, doi: 10.24412/2071-6168-2023-8-296-297.
- [30] Т. Т. З. Нгуен and Л. В. Черненькая, “Фаззификация в моделях прогнозирования нечетких временных рядов,” *Журнал Известия Тульского государственного университета – Технические науки (ТулГУ, г. Тула)*, vol. 8, no. Системный анализ, Управление и обработка информации, pp. 337–346, 2023.
- [31] Т. Т. З. Нгуен and Л. В. Черненькая, “Модель Прогнозирования Эвристических Нечетких Временных Рядов Высокого Порядка, Основанная На Хедж-Алгебраическом Подходе Часть 2,” *Журнал Известия Тульского государственного университета – Технические науки (ТулГУ, г. Тула)*, vol. 9, no. Системный анализ, Управление и обработка информации, 2023.
- [32] Т. Т. З. Нгуен and Л. В. Черненькая, “Модель Прогнозирования Эвристических Нечетких Временных Рядов Высокого Порядка, Основанная На Хедж-Алгебраическом Подходе Часть 3,” *Журнал Известия Тульского государственного университета – Технические науки (ТулГУ, г. Тула)*, vol. 9, no. Системный анализ, Управление и обработка информации, 2023.
- [33] Нгуен Тхи Тху Зунг and Васильевна Черненькая Людмила, “Модель Прогнозирования Эвристических Нечетких Временных Рядов Высокого Порядка, Основанная На Хедж-Алгебраическом Подходе Часть 1,” *Журнал Известия Тульского государственного университета – Технические науки (ТулГУ, г. Тула)*, vol. 9, no. Системный анализ, Управление и обработка информации, 2023.
- [34] T. T. D. Nguyen and L. V. Chernenkaya, “Forecasting model of intuitionistic fuzzy time series using ratio distribution,” *International Journal of Open Information Technologies*, vol. 11, no. 11, pp. 35–44, 2023.
- [35] J. Li, L. Yang, Y. Qu, and G. Sexton, “An extended Takagi–Sugeno–Kang inference system (TSK+) with fuzzy interpolation and its rule base generation,” *Soft comput*, vol. 22, no. 10, pp. 3155–3170, 2018, doi: 10.1007/s00500-017-2925-8.
- [36] M. A.-H. Basil, J. Agustin, and M. Fernando, “A new approach to fuzzy estimation of Takagi-Sugeno model and its applications to optimal control for nonlinear systems,” *Applied Soft Computing Journal*, vol. 12, no. 1, pp. 280–290.
- [37] H. Mohd Pauzi and L. Abdullah, “Intuitionistic fuzzy inference system with weighted comprehensive evaluation considering standard deviation-cosine entropy: a fused forecasting model,” *Neural Comput Appl*, vol. 34, no. 14, pp. 11977–11999, 2022, doi: 10.1007/s00521-022-07082-y.
- [38] С. В. И. Калинина В. Н., “Введение в многомерный статистический анализ: Учебное пособие,” 2003.
- [39] А. Ю. Филатов, *Конспект лекций по многомерным статистическим методам: учеб. пособие*, Иркутск: Иркут. ун-Т. 2007.
- [40] T. J. Ross, *Fuzzy logic with engineering applications*. John Wiley, 2000.
- [41] L.-X. Wang and M. J. Mendel, “Generating fuzzy rules by learning from examples,” in *Proceedings of the 1991 IEEE International Symposium on Intelligent Control*, IEEE, Ed., Arlington Virginia, USA: IEEE Xplore, pp. 263–268.

A new Gaussian fuzzy logic inference system of Takagi-Sugeno-Kanga type with principal component weighting

Nguyen Thi Thu Dung

Abstract—At the present time, fuzzy inference systems have given remarkably effective support in solving many problems in practical applications. Among such systems, the powerful characteristics and applications of the Takagi-Sugeno-Kanga (TSK) fuzzy system are significant. In this study, a new Gaussian fuzzy TSK inference system with the use of principal component analysis method is proposed to minimize the volume capacity of fuzzy logic rule system when the number of input indicators is relatively large, and the same time, the model has enhancements when weight values of input indicators are considered by the proportion of input information contribution. The model applies the entropy minimization approach (MEPA) to support the fuzzification process of a number of input data in an efficient way. The proposed model is applied to forecast the socio-economic development index of 63 provinces in Vietnam, which uses the socio-economic development indicators of 63 provinces in 2019 as input data. The forecasting results of the proposed system are measured and analyzed by comparison of forecasted and actual values, and evaluation of MSE, RMSE, MAPE and CORR values between the model using weighting coefficients and the model not using weighting coefficients. The proposed model using weighting factors has better performance

Keywords—Takagi-Sugeno-Kanga fuzzy inference system, entropy minimization principle (MEPA), principal component weights, reduced criteria, rule-based fuzzy system, socio-economic index.

REFERENCES

- [1] L. A. Zadeh, "Fuzzy Sets," *Information and control*, vol. 8, pp. 338–353, 1965.
- [2] O. Castillo and P. Melin, "A review on interval type-2 fuzzy logic applications in intelligent control," *Inf Sci (N Y)*, vol. 279, pp. 615–631, 2014.
- [3] S. H. Liao, "Expert system methodologies and applications - a decade review from 1995 to 2004," *Expert Syst Appl*, vol. 28, no. 1, pp. 93–103, 2005.
- [4] C. C. Lee, "Fuzzy logic in control systems: fuzzy logic controller," *IEEE Trans Syst Man Cybern*, vol. 20, no. 2, pp. 404–418, 1990.
- [5] H. O. Wang, K. Tanaka, and M. F. Griffin, "An approach to fuzzy control of nonlinear systems: Stability and design issues," *IEEE Trans Syst Man Cybern*, vol. 4, no. 1, pp. 14–23, 1996.
- [6] R.-E. Precup and H. Hans, "A survey on industrial applications of fuzzy control," *Comput Ind*, vol. 62, no. 3, pp. 213–226.
- [7] M. Komiyama, K. Yoshimoto, M. Sisido, and K. Ariga, "Chemistry can make strict and fuzzy controls for bio-systems: DNA nanoarchitectonics and cell-macromolecular nanoarchitectonics," *The Chemical Society of Japan*, vol. 90, no. 9, pp. 967–1004, 2017.
- [8] Y. Jin and L. Wang, *Fuzzy Systems in Bioinformatics and Computational biology*, Studies in. Springer, 2009.
- [9] G. Bojadziev, *Fuzzy logic for business, finance, and management*, vol. 23. 2007.
- [10] J. Gomez and D. Dasgupta, "Evolving fuzzy classifiers for intrusion detection," in *2002 IEEE Workshop on inf. Assur*, I. C. Press, Ed., New York, 2002, pp. 321–323.
- [11] S. Elhag, A. Fernandez, A. Bawakid, S. Alshomrani, and F. Herrera, "On the combination of genetic fuzzy systems and pairwise learning for improving detection rates on intrusion detection systems," *Expert Syst Appl*, vol. 42, no. 1, pp. 193–202, 2015.
- [12] A. Lemos, W. Caminhas, and G. Fernando, "Adaptive fault detection and diagnosis using an evolving fuzzy classifier," *Inf Sci (N Y)*, vol. 220, pp. 64–85, 2013.
- [13] P. Melin, O. Mendoza, and O. Castillo, "Face recognition with an improved interval type-2 fuzzy logic sugeno integral and modular neural networks," *IEEE Transactions on Systems, Man and Cybernetics, Part A. Syst. Humans*, vol. 41, no. 5, pp. 1001–1012, 2011.
- [14] E. H. Mamdani, "Application of fuzzy algorithms for control of simple dynamic plant," *Proceedings of the Institution of Electrical Engineers*, vol. 121, no. 12, pp. 1585–1588, 1974.
- [15] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control," *IEEE Trans Syst Man Cybern*, vol. 15, no. 1, pp. 116–132, 1985.
- [16] B. Mohammed and A. Mostafa, "Comparison of Mamdani-type and Sugeno-Type Fuzzy Inference Systems for Fuzzy Real Time Scheduling," *International Journal of Applied Engineering Research*, vol. 11, no. 22, pp. 11071–11075, 2016.
- [17] S. Tina *et al.*, "A comparative Analysis of the Mamdani and Sugeno Fuzzy Inference systems for MPPT of an ISlanded PV System," *Int J Energy Res*, vol. 2023, p. 14, 2023.
- [18] T. Tettey and T. Marwala, "Modelling Conflict: Knowledge Extraction using Bayesian Neural Network and Neuro-fuzzy Models," *Comput Intell*, no. Mid, 2016.
- [19] S. G. Cao, N. W. Rees, and G. Feng, "Analysis and design for a class of complex control systems - Part I: fuzzy modeling and identification, *Automatica*," *Automatica*, vol. 33, no. 6, pp. 1017–1028.
- [20] T. P. Hong and C. Y. Lee, "Induction of fuzzy rules and membership functions from training examples," *Fuzzy Sets Syst*, vol. 84, no. 1, pp. 33–47, 1996, doi: 10.1016/0165-0114(95)00305-3.

- [21] F. Klawonn and R. Kruse, "Constructing a fuzzy controller from data," 1997.
- [22] T. A. Johansen, R. Shorten, and R. Murray-Smith, "On the interpretation and identification of dynamic Takagi-Sugeno fuzzy models," *IEEE Transactions on Fuzzy Systems*, vol. 8, no. 3, pp. 297–313, 2000, doi: 10.1109/91.855918.
- [23] S. Petrovic-Lazarevic, "Neuro-Fuzzy Support of Knowledge Management in Social Regulation," no. February, pp. 387–400, 2003, doi: 10.1063/1.1503710.
- [24] I. Škrjanc, S. Blažič, and O. Agamennoni, "Interval fuzzy model identification using l_∞ -norm," *IEEE Transactions on Fuzzy Systems*, vol. 13, no. 5, pp. 561–568, 2005, doi: 10.1109/TFUZZ.2005.856567.
- [25] M. Aqil, I. Kita, A. Yano, and S. Nishiyama, "A Takagi-Sugeno fuzzy system for the prediction of river stage dynamics," *Jpn Agric Res Q*, vol. 40, no. 4, pp. 369–378, 2006, doi: 10.6090/jarq.40.369.
- [26] K. Simiński, "Rule weights in a neuro-fuzzy system with a hierarchical domain partition," *International Journal of Applied Mathematics and Computer Science*, vol. 20, no. 2, pp. 337–347, 2010, doi: 10.2478/v10006-010-0025-3.
- [27] S. Surono, G. K. Wen, C. W. Onn, Y. Bin Dasril, A. Y. Astuti, and N. Periasamy, "Effectiveness of Enhanced Takagi Sugeno Kang'S Fuzzy Inference Model," *J Theor Appl Inf Technol*, vol. 100, no. 8, pp. 2383–2392, 2022.
- [28] C. T. Lin and G. C. S. Lee, "Neural-Network-Based Fuzzy Logic Control and Decision System," *IEEE Transactions on Computers*, vol. 40, no. 12, pp. 1320–1336, 1991, doi: 10.1109/12.106218.
- [29] Nguyen Thi Thu Dung and L. V. Chernenkaya, "Discretization in fuzzy time series forecasting models," *Journal of Tula State University - Technical Sciences (Tula State University, Tula)*, vol. 8, no. Systems Analysis, Control and Information Processing, pp. 296-304, 2023, doi: 10.24412/2071-6168-2023-8-296-297.
- [30] T. T. D. Nguyen and L. V. Chernenkaya, "Fuzzification in fuzzy time series forecasting models," *Journal of Tula State University - Technical Sciences (TulGU, Tula)*, vol. 8, no. Systems Analysis, Control and Information Processing, pp. 337-346, 2023.
- [31] T. T. D. Nguyen and L. V. Chernenkaya, "Heuristic Fuzzy High Order Time Series Forecasting Model Based on Hedge-Algebraic Approach Part 2," *Journal of Tula State University - Technical Sciences (TulGU, Tula)*, vol. 9, no. System Analysis, Control and Information Processing, 2023.
- [32] T. T. D. Nguyen and L. V. Chernenkaya, "Heuristic Fuzzy High Order Time Series Forecasting Model Based on Hedge-Algebraic Approach Part 3," *Journal of Tula State University - Technical Sciences (TulGU, Tula)*, vol. 9, no. System Analysis, Control and Information Processing, 2023.
- [33] Nguyen Thi Thu Dung and Vasilievna Chernenkaya Lyudmila, "Heuristic Fuzzy High Order Time Series Forecasting Model Based on Hedge-Algebraic Approach Part 1," *Journal of Tula State University - Technical Sciences (Tula State University, Tula)*, vol. 9, no. System Analysis, Control and Information Processing, 2023.
- [34] T. T. D. Nguyen and L. V. Chernenkaya, "Forecasting model of intuitionistic fuzzy time series using ratio distribution," *International Journal of Open Information Technologies*, vol. 11, no. 11, pp. 35–44, 2023.
- [35] J. Li, L. Yang, Y. Qu, and G. Sexton, "An extended Takagi–Sugeno–Kang inference system (TSK+) with fuzzy interpolation and its rule base generation," *Soft comput*, vol. 22, no. 10, pp. 3155–3170, 2018, doi: 10.1007/s00500-017-2925-8.
- [36] M. A.-H. Basil, J. Agustin, and M. Fernando, "A new approach to fuzzy estimation of Takagi-Sugeno model and its applications to optimal control for nonlinear systems," *Applied Soft Computing Journal*, vol. 12, no. 1, pp. 280–290.
- [37] H. Mohd Pauzi and L. Abdullah, "Intuitionistic fuzzy inference system with weighted comprehensive evaluation considering standard deviation-cosine entropy: a fused forecasting model," *Neural Comput Appl*, vol. 34, no. 14, pp. 11977–11999, 2022, doi: 10.1007/s00521-022-07082-y.
- [38] C. V. I. Kalinina V. N., "Introduction to Multivariate Statistical Analysis: Textbook," 2003. (In Russian)
- [39] A. Yu. Filatov, *Lecture notes on multivariate statistical methods: textbook*, Irkutsk: Irkut. unT. 2007. (In Russian)
- [40] T. J. Ross, *Fuzzy logic with engineering applications*. John Wiley, 2000.
- [41] L.-X. Wang and M. J. Mendel, "Generating fuzzy rules by learning from examples," in *Proceedings of the 1991 IEEE International Symposium on Intelligent Control*, IEEE, Ed., Arlington Virginia, USA: IEEE Xplore, pp. 263–268.