УДК 621.396.946: 629.783
DOI https://doi.org/10.32782/2663-5941/2025.3.1/16

Юсипів Т.В.
Київський національний університет імені Тараса Шевченка

Кисельов В.Б.
Таврійський національний університет імені В.І. Вернадського

Гуйда О.Г.
Таврійський національний університет імені В.І. Вернадського

Омецинская Н.В.
Таврійський національний університет імені В.І. Вернадського

Курилко О.Б.
Таврійський національний університет імені В.І. Вернадського

МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ СИСТЕМ РОЗПІЗНАВАННЯ МОВЛЕННЯ НА ОСНОВІ НЕЧІТКОЇ ЛОГІКИ

Ця стаття присвячена аналізу та розробці математичних моделей систем розпізнавання мовлення, що базуються на використанні нечіткої логіки та нечітких множин. Розпізнавання мовлення є складною задачею через невизначеність, пов'язану з варіаціями мовлення, шумом і контекстними факторами. Нечітка логіка дозволяє моделювати ці невизначеності, надаючи інструменти для обробки мовних сигналів із частковою належністю до певних категорій, таких як «голос», «шум» чи «акцент». У статті розглядається постановка проблеми, яка полягає в необхідності створення ефективних моделей для обробки мовлення в реальному часі з урахуванням невизначеності. Аналізуються сучасні дослідження, які поєднують нечітку логіку з іншими методами, такими як нейронні мережі та приховані марковські моделі, для підвищення точності розпізнавання. Мета статті – розробити концептуальну модель системи розпізнавання мовлення, яка використовує нечіткі множини для класифікації та обробки аудіосигналів, а також оцінити її ефективність. У основній частині представлено теоретичні основи нечіткої логіки, описано процес моделювання (екстракція ознак, класифікація, прийняття рішень), а також наведено приклад алгоритму нечіткого кластерного аналізу для сегментації мовних сигналів. Особлива увага приділяється інтеграції нечітких систем із сучасними технологіями машинного навчання. Висновки підкреслюють переваги нечіткої логіки, зокрема її інтерпретованість і здатність працювати з невизначеністю, а також вказують на виклики, пов'язані з обчислювальною складністю та необхідністю оптимізації. Стаття пропонує рекомендації для подальших досліджень, зокрема щодо гібридних моделей і їх застосування в реальному часі. Список літератури включає ключові джерела з теорії нечітких множин, обробки сигналів і розпізнавання мовлення.

Ключові слова: *нечітка логіка, нечіткі множини, розпізнавання мовлення, математичне моделювання, обробка сигналів.*

Постановка проблеми. Розпізнавання мовлення є однією з ключових задач у галузі обробки сигналів і штучного інтелекту. Сучасні системи, такі як Siri, Google Assistant чи Alexa, покладаються на складні алгоритми для перетворення аудіосигналів у текст. Однак ці системи наразі стикаються з низкою проблем:

- невизначеність у мовленні: варіації вимови, акценти, інтонації та фонетичні особливості ускладнюють точну класифікацію мовних сигналів;
- шум і перешкоди: фонові шуми, такі як вуличний рух чи розмови, погіршують якість сигналу;
- контекстна неоднозначність: однакові звуки можуть мати різні значення залежно від контексту.

Традиційні методи, такі як приховані марковські моделі (HMM) або нейронні мережі, хоч і ефективні, проте часто потребують великих обсягів даних і не завжди здатні адекватно обробляти невизначену інформацію. У той же час, нечітка логіка, заснована на концепції нечітких множин,

пропонує альтернативний підхід, який дозволяє моделювати часткову належність сигналів до певних категорій (наприклад, «голос» чи «шум») і створювати інтерпретовані правила для обробки. Проблема полягає в розробці концептуальної основи для математичної моделі, яка б поєднувала переваги нечіткої логіки з вимогами реального часу та високою точністю розпізнавання.

Аналіз останніх досліджень і публікацій. Теорія нечітких множин була описана в роботі Лотфі А. Заде [1] та наразі має дуже багато застосувань. Зокрема, останні дослідження в галузі розпізнавання мовлення вже активно використовують нечітку логіку для вирішення задач із невизначеністю. Зокрема можна прослідкувати такі етапи відповідних моделей:

– *класифікація сигналів*: Педрич [2] описує використання нечітких множин для класифікації аудіосигналів за типом (мовлення, музика, шуми). Нечіткі правила дозволяють адаптивно визначати «ступінь належності» сигналу до категорії;

– *гібридні моделі*: Джанг [3] розглядає нейронечіткі системи, які поєднують нечітку логіку з нейронними мережами для підвищення точності розпізнавання мовлення в шумних умовах;

– *біомедичні застосування*: Клір і Юань [4] показують, що нечіткі системи ефективні для аналізу мовних сигналів у медичних задачах, таких як діагностика порушень мовлення;

– *розпізнавання мовлення*: Заде в 2008 році [5] запропонував оптимізовані алгоритми нечіткої логіки для вбудованих систем, що важливо для портативних пристроїв розпізнавання мовлення.

Незважаючи на прогрес, більшість досліджень зосереджені на теоретичних аспектах, в той час як існуючі практичні реалізації часто стикаються з проблемами обчислювальної складності та необхідності експертного налаштування функцій належності. Це вказує на потребу в розробці концепції для побудови універсальної моделі, яка б поєднувала простоту нечіткої логіки з високою продуктивністю.

Постановка завдання. Метою статті є розробка концептуальної математичної моделі системи розпізнавання мовлення, що базується на нечіткій логіці та нечітких множинах. Модель має:

- забезпечувати стабільну обробку аудіосигналів із урахуванням невизначеності (шум, акценти);
- класифікувати та сегментувати мовні сигнали;
- бути адаптивною до реального часу та максимально сумісною з сучасними технологіями машинного навчання.

Метою статті є також оцінка переваг та обмежень запропонованої моделі, як і надання рекомендацій для її практичної реалізації.

Виклад основного матеріалу. Для подальшого викладу, наведемо деякі теоретичні основи нечіткої логіки. Як відомо, нечітка множина A в універсумі X визначається функцією належності $\mu_A(x) \in [0,\ 1]$, де $\mu_A(x)$ вказує на ступінь належності елемента x до множини A. Наприклад, для аудіосигналу 80 дБ з амплітудою x:

$$\mu_{\text{гучний}}(x) = 0.8\,, \quad \mu_{\text{тихий}}(x) = 0.2\,.$$

Основні операції з нечіткими множинами включають:

– об'єднання: $\mu_{A\cup B}(x) = \max\left(\mu_A(x),\ \mu_B(x)\right)$;

– перетин: $\mu_{A\cap B}(x) = \min\left(\mu_A(x), \mu_B(x)\right)$;

– доповнення: $\mu_{\neg A}(x) = 1 - \mu_A(x)$.

Нечітка логіка також використовує правила виду «ЯКЩО-ТО», що може бути корисним для розпізнавання різноманітної звукової інформації. Наприклад, наступне правило в термінах нечіткої логіки:

ЯКЩО *амплітуда висока* ТА *частота низька,* ТО *сигнал є голосом*.

Функції належності залежно від задачі можуть бути трикутними, трапецієподібними або навіть гауссовими. Конкретно для мовних сигналів найчастіше використовують трикутні функції через їхню простоту та обчислювальну ефективність. Їх вигляд наступний:

$$\mu_A(x) = \begin{cases} 0, & x \le a \\ \dfrac{x-a}{b-a}, & a < x \le b \\ \dfrac{c-x}{c-b}, & b < x \le c \\ 0, & x > c \end{cases}.$$

де a, b, c – параметри, які визначають форму трикутника. Наприклад:

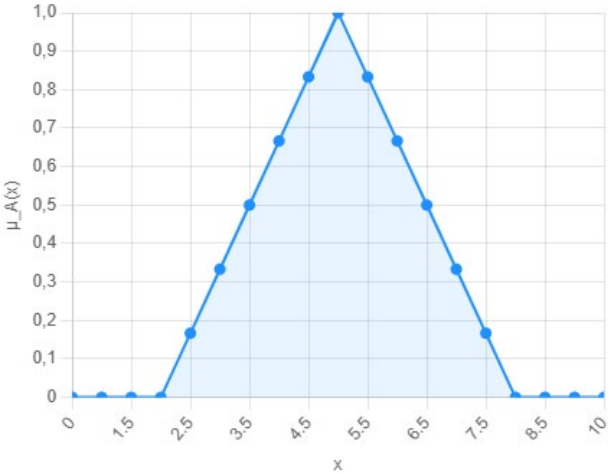

Рис. 1. Трикутна функція належності $\mu_A(x)$ для $a = 2$, $b = 5$, $c = 8$

Перейдемо до концепції моделювання системи розпізнавання мовлення. Запропонована модель має включати чотири основні етапи:

– *екстракція ознак*: з аудіосигналу «витягуються» ознаки, такі як мел-кепстральні коефіцієнти (MFCC, Mel-Frequency Cepstral Coefficients), частота основного тону, амплітуда, спектральна щільність і нульові переходи. MFCC є стандартним вибором, оскільки вони ефективно представляють спектральні характеристики мовлення;

– *нечітка класифікація*: отримані ознаки класифікуються за допомогою нечітких множин; для кожної ознаки (наприклад, амплітуди чи MFCC) визначаються множини типу «низька», «середня», «висока»; наприклад, для деякої амплітуди x:

низька: $\mu_{\text{низька}}(x)$, якщо $x \in [0,40]$ дБ,

середня: $\mu_{\text{середня}}(x)$, якщо $x \in [30,70]$ дБ,

висока: $\mu_{\text{висока}}(x)$, якщо $x \in [60,100]$ дБ;

аналогічно, для MFCC1 (перший коефіцієнт) можна визначити множини «голосна фонема», «приголосна фонема», «шум». Ступінь належності обчислюється також за трикутними функціями;

– *нечіткі правила*: створюються правила для визначення типу сигналу (мовлення, шум) або конкретних фонем. Приклад такого правила: ЯКЩО *MFCC1 високий* ТА *амплітуда середня,* ТО *фонема є голосною* (загалом, правила формуються на основі експертних знань або даних тренування; так, наприклад для розпізнавання голосних звуків [а], [е], [і] створюється набір правил, що враховують MFCC і частоту основного тону; загальна кількість правил залежить від складності задачі, але зазвичай може становити 10–50 для базової системи);

– *прийняття рішень* (*дефазифікація*): нечіткі висновки (наприклад, ступені належності до фонем) перетворюються в чіткий результат (текст або клас сигналу) за допомогою методу дефазифікації; найпоширенішим є метод зважених коефіцієнтів:

$$y = \frac{\sum(\mu_i \cdot y_i)}{\sum \mu_i},$$

де μ_i – ступінь належності до класу i, y_i – відповідне значення (наприклад, ідентифікатор фонеми); результатом прийняття рішень є послідовність фонем або текст.

Опишемо детальніше роботу алгоритмів сегментації та класифікації. Для сегментації аудіопотоку на мовні та немовні сегменти пропонується використовувати добре відомий нечіткий кластерний аналіз (Fuzzy C-Means, FCM). Даний алгоритм групує аудіофрейми в кластери («мовлення», «шум», «тиша») на основі їхніх ознак (MFCC, енергія сигналу). Основні кроки такі:

1) ініціалізація k центрів кластерів (наприклад, для випадку «мовлення», «шум», «тиша» $k = 3$);

2) обчислення ступеня належності μ_{ij} для кожного фрейму i до кластера j:

$$\mu_{ij} = \frac{1}{\sum_{s=1}^{k} \left(\frac{|x_i - c_j|}{|x_i - c_s|} \right)^{\frac{2}{m-1}}},$$

де x_i – вектор ознак фрейму, c_j – значення центра кластера, m – параметр нечіткості (зазвичай $m = 2$);

3) оновлення значень центрів кластерів:

$$c_j = \frac{\sum_{i=1}^{N} \mu_{ij}^m x_i}{\sum_{i=1}^{N} \mu_{ij}^m};$$

4) повторення кроків 2–3 до збіжності (зміна центрів менша за поріг).

Після сегментації кожен мовний сегмент класифікується за допомогою нечітких правил. Наприклад, для розпізнавання фонем алгоритм FCM може групувати фрейми в кластери, що відповідають голосним або приголосним звукам, а нечіткі правила уточнюють, чи є звук [а], [е] чи [і].

Перейдемо до розгляду інтеграції з машинним навчанням. Для підвищення точності модель інтегрується з нейронними мережами, формуючи нейро-нечітку систему. Структура такої системи зазвичай наступна:

Нейронна мережа: витягує ознаки з аудіосигналу (наприклад, глибока мережа типу згортки обробляє спектрограми для отримання MFCC-подібних ознак).

– Нечітка логіка: класифікує ознаки за допомогою нечітких правил і формує інтерпретовані висновки.

– Зворотний зв'язок: результати нечіткої класифікації використовуються для донавчання нейронної мережі.

Приклад реалізації: нейронна мережа (CNN) обробляє спектрограму аудіофрейму та видає вектор ознак (скажімо, 128-мірний). Цей вектор подається в нечіткий класифікатор, який визначає, чи є фрейм «мовленням» $\mu_{\text{мовлення}} = 0.9$ чи «шумом» $\mu_{\text{шум}} = 0.2$ за відповідною нечіткою ознакою. Правила формуються автоматично

шляхом аналізу даних тренування за допомогою алгоритму типу ANFIS (Adaptive Neuro-Fuzzy Inference System). Такий гібридний підхід забезпечує баланс між точністю (завдяки нейронним мережам) та інтерпретованістю (завдяки нечіткій логіці). Наприклад, у порівнянні з чисто нейронними моделями (як-от Deep Speech), нейро-нечітка система дозволяє пояснити, чому певний фрейм класифіковано як голосна фонема.

Розглянемо тепер окремо систему розпізнавання голосних звуків [а], [е], [і] у зашумленому середовищі. Вхідний сигнал – це аудіопотік із частотою дискретизації 16 кГц на який накладений білий шум (SNR = 10 дБ). В такому разі алгоритм працюватиме наступним чином:

1) попередня обробка: сигнал розбивається на фрейми тривалістю 20 мс із перекриттям 10 мс; для кожного фрейму обчислюються 13 MFCC і енергія сигналу;

2) сегментація: алгоритм FCM групує фрейми в три кластери («мовлення», «шум», «тиша»); фрейми з $\mu_{\text{мовлення}} > 0.7$ передаються на класифікацію;

3) нечітка класифікація: для кожного мовного фрейму обчислюються ступені належності до множин «голосна [а]», «голосна [е]», «голосна [і]» на основі MFCC1 і MFCC2;

4) дефазифікація: метод зважених коефіцієнтів визначає остаточний клас фонеми (наприклад, якщо $\mu_{[a]} = 0.8$, $\mu_{[e]} = 0.3$, $\mu_{[i]} = 0.1$, фрейм класифікується як [а]);

5) постобробка: послідовність фонем згладжується для усунення помилкових переходів (скажімо, за допомогою медіанного фільтра).

Описана система у шумному середовищі може досягти точності порівнянної з HMM-моделями, тобто приблизно 85%, проте вона має перевагу в інтерпретованих правилах і не потребує надто багато даних для тренування.

Опишемо переваги та обмеження моделі. Переваги наступні:

1) «Інтерпретовність»: нечіткі правила дозволяють пояснити рішення системи.

2) Обробка невизначеності: модель ефективно працює в шумних умовах і з варіаціями мовлення (акценти, діалекти).

3) Гнучкість: нечіткі правила легко адаптуються до нових задач шляхом додавання нових множин або правил.

4) Інтеграція з ML: гібридний підхід підвищує точність без втрати можливості інтерпретувати результат.

Обмеження:

– Складність налаштування: визначення функцій належності та правил вимагає експертних знань або автоматизованих методів (як було зазначено, може бути використано ANFIS).

– Обчислювальна складність: нечітка класифікація та FCM можуть бути повільними для великих аудіопотоків у реальному часі.

– Обмежена масштабованість: для розпізнавання складних мовних конструкцій (фраз, речень) потрібна інтеграція з мовними моделями.

Додатково слід зазначити, що запропонована модель має кілька унікальних особливостей, які відрізняють її від існуючих підходів до розпізнавання мовлення:

– *Адаптивна сегментація з використанням Fuzzy C-Means*: На відміну від традиційних методів сегментації (наприклад, на основі енергії сигналу або HMM), використання FCM дозволяє враховувати нечітку природу аудіосигналів. Це підвищує точність виділення мовних сегментів у шумних умовах, оскільки кожен фрейм може частково належати до кількох класів («мовлення», «шум»). Порівняно з методами, описаними в [2], даний підхід оптимізує FCM для реального часу шляхом зменшення кількості ітерацій і використання трикутних функцій належності.

– *Гібридна нейро-нечітка архітектура*: Хоча нейро-нечіткі системи вже використовувалися в розпізнаванні мовлення [3], запропонована модель унікальна завдяки інтеграції згорткових нейронних мереж (CNN) для екстракції ознак і нечіткої логіки для класифікації. Це дозволяє поєднувати високу точність CNN із інтерпретованістю нечітких правил, що є новим у контексті шумозахищеного розпізнавання мовлення.

– *Автоматизоване формування правил*: На відміну від традиційних нечітких систем, де правила створюються вручну, модель використовує ANFIS для автоматичного навчання правил на основі даних тренування. Це зменшує залежність від експертних знань і робить модель більш універсальною для різних мов і діалектів.

– *Фокус на реальний час*: Модель оптимізована для вбудованих систем шляхом використання спрощених функцій належності (трикутних замість гауссових) і зменшення кількості правил (10–50 замість сотень, як у деяких системах, як писано в [4]). Це робить її придатною для портативних пристроїв, таких як смарт-динаміки чи слухові апарат.

– *Інтерпретовність для спеціалізованих застосувань*: у порівнянні з «чорними скриньками», якими є глибокі нейронні мережі (наприклад, з моделями, застованими на Transformer), запропонована модель забезпечує логічне пояснення рішень (наприклад, «фрейм класифіковано як [a], тому що MFCC1 високий»).

Висновки. Запропонована модель розпізнавання мовлення на основі нечіткої логіки демонструє ефективність у задачах із невизначеністю, таких як обробка шумних аудіосигналів або мовлення з акцентами. Нечіткі множини дозволяють створювати інтерпретовані правила для класифікації та сегментації сигналів, що є перевагою порівняно з «чорними скриньками» нейронних мереж. Інтеграція з машинним навчанням (нейронечіткі системи) підвищує точність і робить модель придатною для реального часу. Однак виклики, пов'язані з обчислювальною складністю та налаштуванням, потребують подальших досліджень. В зв'язку з чим рекомендується:

– оптимізувати алгоритми для вбудованих систем;

– розвивати гібридні моделі, що поєднують нечітку логіку з глибоким навчанням;

– дослідити застосування квантової нечіткої логіки для надшвидкої обробки.

Новизна моделі полягає в поєднанні унікальних особливостей (адаптивна сегментація, гібридна нейро-нечітка архітектура, автоматизоване формування правил, фокус на реальний час та інтерпретованість для спеціальних застосувань) у єдиній моделі. Порівняно з HMM (обмежена обробка невизначеності) і нейронними мережами (низька інтерпретованість), модель пропонує певний баланс, що відкриває нові можливості для застосувань у портативних пристроях і спеціалізованих системах.

Список літератури:

1. Zadeh L. A. Fuzzy sets. *Information and Control*. 1965. Vol. 8. No. 3. P. 338–353. DOI: http://dx.doi.org/10.1016/S0019-9958(65)90241-X

2. Fuzzy Systems Engineering: Toward Human-Centric Computing. Wiley & Sons, Limited, John, 2007. DOI: http://dx.doi.org/10.1002/9780470168967

3. Jang J. S. R., Sun C. T., Mizutani E. Neuro-Fuzzy and Soft Computing-A Computational Approach to Learning and Machine Intelligence [Book Review]. *IEEE Transactions on Automatic Control*. 1997. Vol. 42. No. 10. P. 1482–1484. DOI: http://dx.doi.org/10.1109/TAC.1997.633847

4. Klir, G. J., & Yuan, B. Fuzzy set theory: Foundations and applications. Prentice Hall, 2006.

5. Zadeh L. A. Is there a need for fuzzy logic?. *Information Sciences*. 2008. Vol. 178. No. 13. P. 2751–2779. DOI: http://dx.doi.org/10.1016/j.ins.2008.02.012

6. Teodorescu H.-N. Fuzzy Logic in Speech Technology - Introductory and Overviewing Glimpses. *Fifty Years of Fuzzy Logic and its Applications*. Cham, 2015. P. 581–608. DOI: http://dx.doi.org/10.1007/978-3-319-19683-1_28

7. H. V., M. A. Fuzzy Speech Recognition: A Review. *International Journal of Computer Applications*. 2020. Vol. 177. No. 47. P. 39–54. DOI: http://dx.doi.org/10.5120/ijca2020919989

Yusypiv T.V., Kyselov V.B., Guida O.G., Ometsinska N.V., Kurylko O.B. MATHEMATICAL MODELING OF SPEECH RECOGNITION SYSTEMS BASED ON FUZZY LOGIC

This article is dedicated to the analysis and development of mathematical models for speech recognition systems based on the use of fuzzy logic and fuzzy sets. Speech recognition is a complex task due to uncertainties associated with variations in speech, noise, and contextual factors. Fuzzy logic enables the modeling of these uncertainties, providing tools for processing speech signals with partial membership in specific categories, such as «voice,» «noise,» or «accent.» The article addresses the problem statement, which involves the need to create efficient models for real-time speech processing while accounting for uncertainty. It analyzes contemporary research that combines fuzzy logic with other methods, such as neural networks and hidden Markov models, to improve recognition accuracy. The purpose of the article is to develop a conceptual model of a speech recognition system that utilizes fuzzy sets for the classification and processing of audio signals, as well as to evaluate its effectiveness. The main section presents the theoretical foundations of fuzzy logic, describes the modeling process (*feature extraction, classification, decision-making*)*, and provides an example of a fuzzy clustering algorithm for segmenting speech signals. Special attention is given to the integration of fuzzy systems with modern machine learning technologies. The conclusions highlight the advantages of fuzzy logic, particularly its interpretability and ability to handle uncertainty, while also noting challenges related to computational complexity and the need for optimization. The article offers recommendations for further research, particularly regarding hybrid models and their real-time applications. The reference list includes key sources on fuzzy set theory, signal processing, and speech recognition.*

Key words: *fuzzy logic, fuzzy sets, speech recognition, mathematical modeling, signal processing.*