DOI: https://doi.org/10.36910/6775-2524-0560-2024-55-09
УДК 004.02
Дідус Андрій Володимирович, аспірант
https://orcid.org/0009-0004-2235-6742
Терейковський Ігор Анатолійович, д.т.н., професор
https://orcid.org/0000-0003-4621-9668
Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», м. Київ, Україна

ФОРМАЛІЗАЦІЯ ПРОЦЕСУ РОЗПІЗНАВАННЯ КЛЮЧОВИХ СЛІВ У ГОЛОСОВОМУ СИГНАЛІ

Дідус А. В., Терейковський І. А. Формалізація процесу розпізнавання ключових слів у голосовому сигналі. У даній роботі було здійснено формалізацію процесу розпізнавання ключових слів у голосовому сигналі, що стало основою для розробки деталізованої концептуальної моделі. Запропонована модель надає комплексний і систематизований опис процесу створення ефективних засобів розпізнавання. Модель охоплює критичні компоненти такі як виділення ознак, методологію моделювання, стратегію навчання, процедури тестування, та механізми використання. Значна увага приділяється впровадженню передових методів глибокого навчання, включаючи нейронні мережі, марковські моделі та техніки аугментації даних, що сприяє підвищенню точності розпізнавання ключових слів. Ця робота враховує численні виклики, пов'язані з варіативністю голосових сигналів, стійкістю до шумів, обмеженістю навчальних даних та вимогами до продуктивності в режимі реального часу. Запропонований формалізований підхід дозволяє оптимізувати взаємодію між компонентами та підвищити загальну ефективність системи розпізнавання ключових слів у голосовому сигналі.
Ключові слова: розпізнавання ключових слів, формалізація процесу, голосовий сигнал, концептуальна модель, глибоке навчання, нейронні мережі, марковські моделі

Didus A., Tereikovskyi I. Formalization of the keyword recognition process in speech signal. In this work, the formalization of the keyword recognition process in speech signals has been carried out, which has laid the foundation for the development of a detailed conceptual model. The proposed model provides a comprehensive and systematic description of the process for creating effective recognition tools. The model covers critical components such as feature extraction, modeling methodology, training strategy, testing procedures, and usage mechanisms. Significant attention is given to the implementation of advanced deep learning methods, including neural networks, Markov models, and data augmentation techniques, which contribute to enhancing the accuracy of keyword recognition. This work takes into account numerous challenges associated with the variability of speech signals, noise resilience, limited training data, and real-time performance requirements. The proposed formalized approach allows for the optimization of interaction among components and enhances the overall efficiency of the keyword recognition system in speech signals.
Keywords: keyword recognition, process formalization, speech signal, conceptual model, deep learning, neural networks, Markov models.

Постановка наукової проблеми.

В сучасних умовах ефективне розпізнавання ключових слів у голосовому сигналі є важливою та актуальною задачею в області обробки природної мови та голосових технологій. Застосування засобів розпізнавання ключових слів у голосовому сигналі знаходить широке застосування у різноманітних галузях, таких як голосове керування пристроями, системи розпізнавання мови, інтерфейси "людина-пристрій" тощо.

В останні роки при розробці таких систем активно використовуються найбільш сучасні досягнення в галузі глибокого навчання, зокрема нейронні мережі, марковські моделі та різного виду технології обробки даних, що значно підвищує точність та адаптивність систем до різноманітних умов використання.

В контексті швидкого розвитку глибинного навчання, ця задача залучає значну академічну та комерційну увагу, оскільки відкриває нові можливості для розуміння та взаємодії з голосовими даними у різних сферах. Ця наукова робота присвячена формалізації процесу розпізнавання ключових слів у голосовому сигналі, з акцентом на методах глибинного навчання та алгоритмічних підходах, що стає особливо важливим у сучасних умовах розвитку технологій. Важливість теми підтверджується результатами досліджень [1, 4, 5], де вказано, що розв'язання проблеми ефективного розпізнавання ключових слів у голосовому сигналі належить до актуальних завдань у галузі обробки мовленнєвих даних, розробки систем розпізнавання мовлення та створення голосових інтерфейсів. Дана проблема безпосередньо пов'язана з державними науковими та практичними програмами, спрямованими на розвиток сучасних інформаційних технологій, забезпечення кібербезпеки та захисту інформації в умовах зростаючих обсягів даних та ризиків кібератак. [2, 3]

Аналіз досліджень.

В сучасних науково-прикладних роботах дотичних до теми даного наукового дослідження розглядаються різні підходи до розпізнавання ключових слів з використанням прихованих марківських моделей (HMM), згорткових нейронних мереж (CNN), рекурентних нейронних мереж (RNN), трансформерів, багатозадачного навчання та інших сучасних методів машинного навчання. Зокрема:

1. Застосування HMM в поєднанні з мел-частотними кепстральними коефіцієнтами (MFCC) для виділення ознак голосового сигналу та побудови акустичних моделей є класичним підходом [5].
2. Використання CNN, RNN, їх комбінацій та архітектур на кшталт ResNet, TransformerNet для кінцевого розпізнавання ключових слів без проміжної генерації фонем [8].
3. Багатозадачне навчання для одночасного розпізнавання ключових слів та визначення, чи належить голос користувачеві пристрою чи сторонній особі [7].
4. Застосування методів аугментації даних, зокрема голосової конверсії, для розширення навчальних вибірок [7].
5. Дослідження аудіо-візуальних підходів із використанням відеопотоку артикуляції губ для покращення розпізнавання [8].

В результаті проведеного аналізу науково-практичних робіт в області розпізнавання ключових слів у голосовому сигналі можливо стверджувати про відсутність формалізованого та детального опису процесу розробки засобів розпізнавання ключових слів у голосовому сигналі, що ускладнює можливість визначення перспективних шляхів підвищення їх ефективності.

Адже це могло б відкрити широкий спектр подальших досліджень і розробок. Важливо розробити концептуальну модель підходу до розпізнавання ключових слів у голосовому сигналі, яка б враховувала низку викликів та обмежень, зокрема:

1. Виклик розпізнавання в умовах шуму: Голосові сигнали часто змішуються з фоновими шумами, що ускладнює процес розпізнавання. Необхідно розробити методи, які ефективно враховують цей аспект.
2. Варіативність голосових сигналів: Голосові сигнали можуть значно відрізнятися від особи до особи через вікові, статеві, етнічні та інші характеристики. Модель повинна бути здатна адаптуватися до цієї варіативності.
3. Обмеженість даних для навчання: Для ефективного навчання моделей глибокого навчання потрібні великі набори даних. Однак, в реальному світі, доступ до великих наборів даних може бути обмеженим.
4. Виклики реалізації в реальному часі: Для багатьох застосувань, таких як голосове керування пристроями, необхідно, щоб розпізнавання ключових слів відбувалося в реальному часі. Це вимагає оптимізації як самої моделі, так і процесу обробки сигналів.
5. Проблема перенавчання: При використанні складних моделей, таких як глибокі нейронні мережі, існує ризик перенавчання, коли модель стає занадто специфічною для навчального набору даних і втрачає здатність узагальнювати на нових даних.

Незважаючи на значні досягнення в області розпізнавання ключових слів в голосових сигналах (ГС) завдяки застосуванню різноманітних алгоритмів та моделей, у науковій літературі спостерігається недостатність у формалізації самого процесу розпізнавання, що у свою чергу ускладнює визначення перспективних шляхів удосконалення відомих технологій розпізнавання.

У світлі існуючих викликів та обмежень, важливо розробити детальну концептуальну модель для розпізнавання ключових слів у голосових сигналах. Такий формалізований опис повинен охоплювати всі критичні компоненти процесу, включно з виділенням ознак, методами моделювання, стратегіями навчання, процедурами тестування та механізмами використання.

Відповідно до загальноприйнятої методології розробки засобів розпізнавання [11, 12] результатом досліджень в області формалізації процесу розпізнавання ключових стіл має бути концептуальна модель.

Мета роботи.

Метою даної наукової роботи являється розробка концептуальної моделі розпізнавання ключових слів у голосовому сигналі, що за рахунок детального формалізованого опису означено процесу забезпечує базис методології створення ефективних засобів розпізнавання ключових слів, їх використання.

Виклад основного матеріалу й обґрунтування отриманих результатів дослідження.

В загальному випадку концептуальна модель - це абстрактне, логічне уявлення реального процесу або системи, яке визначає основні компоненти, їх взаємозв'язки та принципи функціонування. [11]

Концептуальна модель розпізнавання ключових слів може бути представлена як інтеграція низки взаємопов'язаних понять, які використовуються для опису даної проблематики, разом з їх властивостями та характеристиками. Концептуальна модель відображає основну ідею, спрямовуючу концепцію для систематизації даних. З огляду на необхідність розробки ефективних рішень для розпізнавання емоцій, використання дефініцій із сфери комп'ютерної і програмної інженерії, а також теорії розпізнавання образів є цілком виправданим.[7, 8, 9]

Розглядаючи комплексну оцінку ефективності концептуальної моделі підходу до розпізнавання ключових слів у голосовому сигналі, необхідно звернути увагу на цілий ряд взаємопов'язаних параметрів.[4, 5, 6] Ефективність в даному контексті охоплює не тільки технічні аспекти виконання програмної системи, але й оцінку ефективного використання ресурсів, а також адаптацію системи до заданих стандартів і умов експлуатації. Основні характеристики, які допомагають оцінити цю ефективність, включають:

1. Часова ефективність. Відображає здатність системи швидко реагувати на вхідні сигнали, забезпечуючи високу швидкість обробки і скорочення загального часу виконання задач.
2. Оптимізація ресурсів. Оцінює раціональність використання обчислювальних ресурсів та пам'яті, а також забезпечення енергоефективності системи.
3. Стандартизація та адаптивність. Визначає ступінь відповідності системи встановленим нормам і стандартам, її гнучкість у різних умовах використання та надійність у розпізнаванні ключових слів.

У процесі розвитку та оптимізації концептуальної моделі підходу до розпізнавання ключових слів у голосовому сигналі, ключовим елементом є створення єдиного термінологічного базису. Така стандартизація спрямована на включення в модель останніх досягнень у сфері обробки мовлення та аудіо сигналів, що гарантує високу адаптивність та відповідність до сучасних наукових реалій. З урахуванням різноманітності існуючих підходів та алгоритмів, слід визначити наступні основні поняття:

1. Ключові слова - це певні слова або фрази в голосовому сигналі, які мають важливе значення для певного контексту або завдання та які необхідно ідентифікувати та розпізнавати системою.
2. Голосовий сигнал (ГС) - складний акустичний сигнал, що генерується голосом людини, який може включати мову, інтонацію, емоційні нюанси та інші звукові характеристики. Голосовий сигнал зазвичай має діапазон від 75 Гц до 8000 Гц, що охоплює різноманітність голосових ефектів, від мовленнєвих тонів до емоційного колориту.
3. Мовний сигнал - це підтип голосового сигналу, присвячений передачі виключно мовної інформації. Мовний сигнал зосереджується на використанні мови, що включає слова, фрази та лінгвістичні структури, і зазвичай локалізується в діапазоні від 80 Гц до 2600 Гц, що оптимізовано для чіткого розуміння мовлення.
4. Акустичні характеристики - параметри, що описують звукові атрибути голосового сигналу, включаючи частоту, інтенсивність, тембр та інші звукові властивості.
5. Гаусівські змішані моделі (GMM) - це статистична модель, яка представляє сукупність гаусівських розподілів ймовірностей, кожен з яких відповідає окремій компоненті (або кластеру) в сукупності даних. У контексті обробки мови, GMM часто використовують для моделювання розподілу акустичних характеристик голосу, дозволяючи ефективно розпізнавати мовні фонеми чи інші одиниці.
6. Нейронна мережа (НМ) - система обробки інформації, заснована на принципах організації та функціонування біологічних нейронних мереж, зокрема мозку.
7. Марковські моделі - це клас статистичних моделей, що використовуються для прогнозування послідовності подій, де ймовірність кожної події залежить лише від стану, досягнутого в попередньому події. Приховані марковські моделі (НММ) є розширенням цієї ідеї, де спостережувані події залежать від внутрішніх факторів, що не можуть бути прямо спостережуваними.

8. Модель марковських ланцюгів (n-грами) - це статистичний метод моделювання мови, який передбачає, що ймовірність появи слова залежить лише від n-1 попередніх слів. Ця модель часто використовується для лінгвістичних моделей в розпізнаванні мови, щоб оцінити ймовірності послідовностей слів або фраз.

9. Модель глибокого навчання - спеціалізована архітектура нейронної мережі, що складається з багатьох шарів, здатна виявляти складні шаблони та закономірності у великих наборах даних.

10. Алгоритми машинного навчання - набір методів і технік, які використовуються для навчання моделей на основі даних, щоб здійснювати прогнозування або приймати рішення.

11. Акустична модель (AM) - це один з основних компонентів систем розпізнавання мови. AM відповідає за перетворення аудіосигналу в послідовність фонем або інших найменших одиниць мовлення. Вона використовує статистичні методи, такі як гаусівські змішані моделі (GMM) або глибокі нейронні мережі (DNN), для визначення ймовірності кожної фонеми в аудіосигналі.

12. Лінгвістична модель (LM) - це модель, який визначає ймовірність послідовності слів або фраз в мовленні. Вона використовує статистичні методи, такі як модель марковських ланцюгів (n-грами), для оцінки ймовірності різних словосполучень.

Ці поняття становлять основу для розуміння та розробки ефективних методів розпізнавання ключових слів, дозволяючи моделі адаптуватися до різних умов та потреб використання.

Для оцінювання ефективності процесу розпізнавання ключових слів у голосовому сигналі, важливо використовувати метрики що відображають особливості піддослідного процесу.[6] Ось перелік ключових метрик, які можуть бути застосовані у дослідницькому процесі для повноцінного аналізу ефективності:

1. Точність (Accuracy): визначає відсоток випадків, коли ключове слово було правильно розпізнане.

$$A = \frac{TP + TN}{TP + FN + FP + TN}, \quad (1)$$

де A - точність, *TP* - True Positives (правильно розпізнані ключові слова), *FN* - False Negatives (неправильно пропущені ключові слова), *FP* - False Positives (неправильно розпізнані ключові слова), *TN* - True Negatives (правильно не розпізнані ключові слова).

2. Повнота (Recall): міра того, скільки реальних випадків ключових слів було виявлено.

$$R = \frac{TP}{TP + FN}, \quad (2)$$

де R - повнота, *TP* - правильно розпізнані ключові слова, *FN* - неправильно пропущені ключові слова.

3. Прецизійність (Precision): відсоток випадків, коли виявлене ключове слово у голосовому сигналі, було дійсно ключовим

$$P = \frac{TP}{TP + FP}, \quad (3)$$

де P - повнота, *TP* - правильно розпізнані ключові слова, *FP* - неправильно розпізнані ключові слова.

4. F-міра (F-Measure): Гармонійне середнє між точністю та відновленням, що допомагає збалансувати ці дві метрики.

$$F = 2 \times \frac{P \times R}{P + R}, \quad (4)$$

де *F* - F-міра, *P* - точність виявлення, *R* - повнота.

5. Оцінка помилок (Error Rate): відсоток випадків, коли ключове слово було неправильно розпізнане або пропущене.

$$ER = \frac{FN + FP}{TP + FN + FP + TN}, \quad (5)$$

де *ER* - оцінка помилок, *TP* - True Positives (правильно розпізнані ключові слова), *FN* - False Negatives (неправильно пропущені ключові слова), *FP* - False Positives (неправильно розпізнані ключові слова), *TN* - True Negatives (правильно не розпізнані ключові слова).

6. Час відгуку (Response Time): час, який системі потрібно для розпізнавання ключового слова після його вимови. Цей показник важливий для систем реального часу.

$$RT = t_{end} - t_{start}, \quad (6)$$

де RT - час відгуку системи, t_{start} - момент початку обробки вхідного сигналу, t_{end} - момент завершення розпізнавання ключового слова та надання відповіді.

Використання цих метрик дозволяє провести всебічну оцінку моделі, враховуючи як її здатність правильно розпізнавати ключові слова, так і реагувати на помилкові спрацьовування. Це забезпечує комплексний підхід до оцінки точності та ефективності моделі.

В процесі створення і удосконалення концептуальної моделі підходу до розпізнавання ключових слів у голосовому сигналі необхідно здійснити аналіз та систематизацію компонентів цієї моделі. Цей процес включає визначення основних елементів, їхніх взаємозв'язків та впливу на загальну продуктивність системи. Нижче наведено загальний опис процесу:

1. Попередня обробка голосових даних: цей крок передбачає підготовку вхідних голосових сигналів до подальшої обробки, включаючи очищення від шумів та нормалізацію.
2. Формування параметрів додаткових даних: на цій стадії відбувається визначення та вибір ознак, які будуть використовуватися для навчання моделі.
3. Формування навчальної та тестової вибірки: розробка та структурування наборів даних, які будуть застосовуватися для навчання моделі та оцінювання її ефективності.
4. Вибір та налаштування моделі: вибір конкретної моделі чи набору моделей для розпізнавання ключових слів і налаштування їх параметрів.
5. Оцінювання ефективності моделі: проведення тестування для оцінки точності та надійності обраної моделі в розпізнаванні ключових слів.
6. Розпізнані слова: кінцевий результат процесу, де модель використовується для виявлення та класифікації ключових слів у реальному аудіосигналі.

Зазначені вище етапи відображають послідовність дій, які необхідно виконати для створення та валідації ефективної моделі розпізнавання ключових слів. Кожен крок, від підготовки даних до фінальної оцінки моделі, є важливою частиною процесу, що веде до точного ідентифікування ключових слів з голосових даних. Всі ці компоненти та їх взаємозв'язки детально зображені на рисунку нижче.

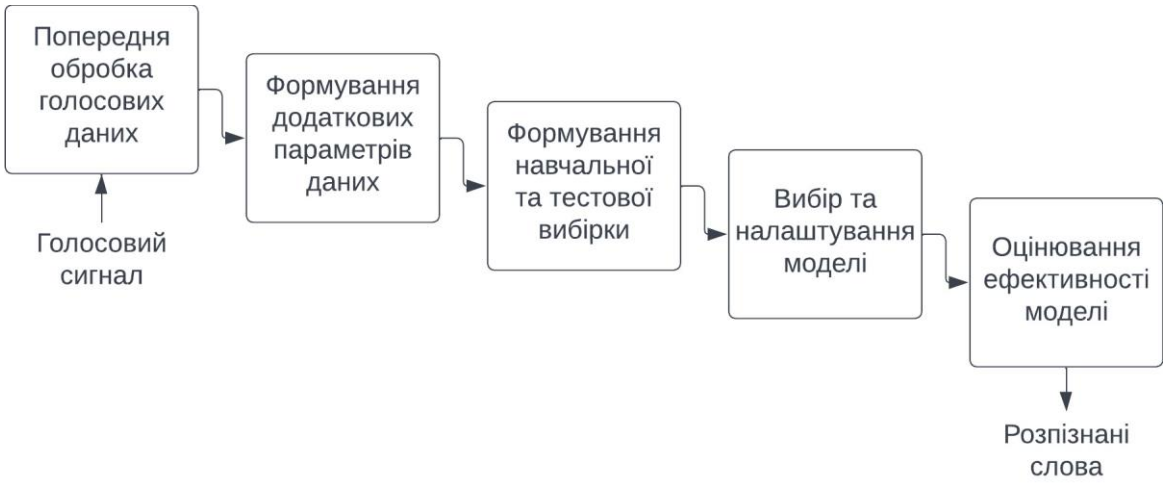

Рис. 1 – Діаграма декомпозиції

Слідом за визначенням основних етапів у створенні концептуальної моделі, було розроблено схему взаємодії компонентів системи розпізнавання ключових слів у голосоваму сигналі, яка показана на рисунку 2. Вона ілюструє методику взаємодії складових у процесі розробки даної системи, формалізуючи взаємодію з важливими складовими, як навчальні та тестові набори даних, вибір оптимальних параметрів моделі та подальшої оцінки її ефективності в контексті розпізнавання ключових слів.

Для розробки схеми взаємодії компонентів концептуальної моделі, яка використовує нейронні мережі, Марковські моделі або алгоритми машинного навчання для розпізнавання ключових слів у голосовому сигналі, можна використовувати таку структуру:

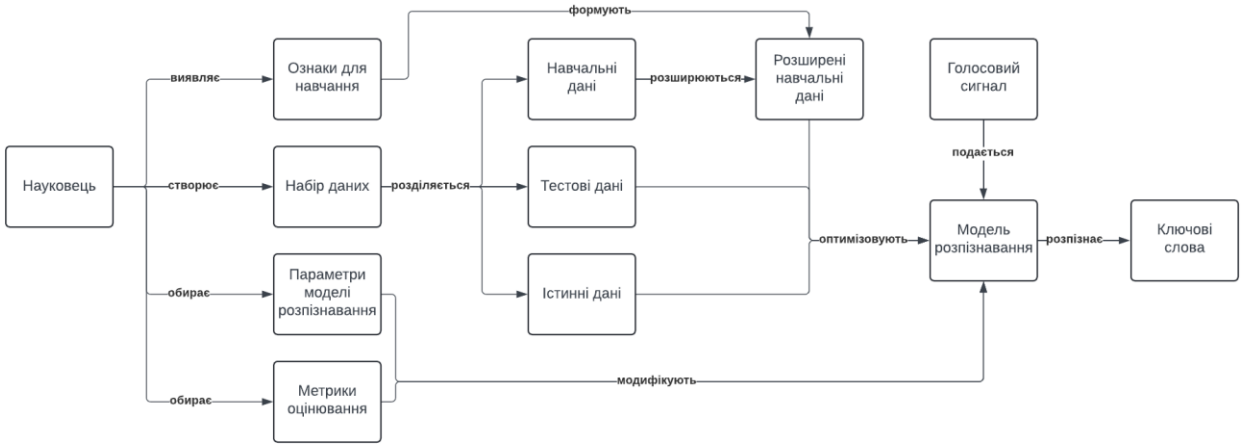

Рис. 2 – Схема взаємодії компонентів системи розпізнавання ключових слів у голосовому сигналі

На схемі, показаній на рисунку 2, детально проілюстровано процес взаємодії елементів системи, починаючи від вибору і формування навчальних даних, до оптимізації параметрів моделювання та врешті - оцінки ефективності моделі. Ці кроки критично важливі для створення ефективного інструменту розпізнавання ключових слів, здатного функціонувати у різноманітних застосуваннях, включаючи, але не обмежуючись, голосовими асистентами та системами автоматичного перекладу. Також у схемі враховано особливості реалізації концептуальної моделі для розпізнавання ключових слів у голосовому сигналі:

- обґрунтування вибору характеристик голосу (ознак для навчання), що використовуються для ідентифікації ключових слів;
- складність визначення оптимального виду моделі та її параметрів для точного розпізнавання ключових слів;
- необхідність удосконалення методів підготовки даних для навчання моделі, яка спрямована на визначення ключових слів.

Таким чином, у розробленій схемі передбачено можливість формування параметрів навчальних даних та їх валідації за допомогою експертної оцінки.

Аналіз структурних компонентів, [12] представлених на попередніх схемах, дає змогу зрозуміти, що ефективність процесу розпізнавання ключових слів у голосовому сигналі залежить від вибору базової моделі розпізнавання, серед яких можуть бути як нейронні мережі, так і інтеграція інших підходів, таких як статистичні методи та Марковські моделі.[9, 10, 11] Ефективність визначених операцій, що включають вибір характеристик голосового сигналу, формування та аналіз навчальних вибірок, а також тонке налаштування параметрів моделей, є вирішальними для точного розпізнавання мовлення. Ця систематизація дає змогу глибше зрозуміти взаємодію між різними компонентами моделі та їх вплив на загальну ефективність системи.

Також важливо уточнити основні операції, які впливають на ефективність розпізнавання емоцій.

1. Підготовка та навчання моделі:

a. Формування навчальних даних: створення репрезентативних наборів даних, які включають різні варіанти ключових слів, їхні контексти та варіації вимови.

b. Розробка навчальної бази даних: збір та систематизація навчальної бази, що охоплює різноманітність мовних та особливостей акцентів.

2. Обробка голосових даних:

a. Вибір ознак для навчання: визначення ключових акустичних та мовних ознак, таких як інтонація, темп, динаміка голосу.

b. Реєстрація та фільтрація даних: запис та відсіювання несуттєвих аспектів голосових сигналів для чіткого визначення ключових слів.

c. Аналіз даних за допомогою нейронних мереж або інших методів: застосування обраних моделей для виявлення та класифікації ключових слів.

3. Проектування та конфігурація моделі:

a. Вибір типу моделі: визначення підходящої архітектури (нейронна мережа, Марковські моделі, комбіновані підходи).

b. Налаштування параметрів моделі: точне налаштування гіперпараметрів для оптимізації точності та ефективності розпізнавання.

Згідно з фундаментальними принципами використання алгоритмів машинного навчання, ефективність систем розпізнавання ключових слів у голосовому сигналі значно зростає завдяки удосконаленню методик аналізу даних. Ці методики застосовуються для оптимізації обробки навчальних вибірок перед самим процесом навчання моделей, призначених для розпізнавання ключових слів.

Отже, формалізовано модель ефективності в контексті розпізнавання ключових слів може бути виражена за допомогою виразу виду:

$E_{total} = f(E_{Dev}, E_{App}, E_{Proc})$, (7)

де E_{total} – інтегральна ефективність системи розпізнавання; E_{Dev}, E_{App} – відповідно, ефективність розробки моделей та їхнього застосування; E_{Proc}– ефективність обробки голосових даних.

В свою чергу кожну зі складових можна деталізувати через залежність від інших складових:

$E_{Dev} = f(w_1, w_2)$; (8)

$E_{App} = f(w_3, w_4)$; (9)

$E_{Proc} = f(w_5, w_6)$, (10)

де w_1- формування навчальних даних: створення репрезентативних наборів даних, які включають різні варіанти ключових слів, їхні контексти та варіації вимови; w_2 - розробка навчальної бази даних: збір та систематизація навчальної бази, що охоплює різноманітність мовних та особливостей акцентів; w_3 - вибір ознак для навчання: визначення ключових акустичних та мовних ознак, таких як інтонація, темп, динаміка голосу; w_4 - реєстрація та фільтрація даних: запис та відсіювання несуттєвих аспектів голосових сигналів для чіткого визначення ключових слів; w_5 - проектування та конфігурація моделі: вибір типу моделі, визначення підходящої архітектури (нейронна мережа, Марковські моделі, комбіновані підходи); w_6 - налаштування параметрів моделі: точне налаштування гіперпараметрів для оптимізації точності та ефективності розпізнавання.

Кожна з цих складових сприяє загальній ефективності системи, забезпечуючи точність, адаптивність та ефективне використання ресурсів у процесі розпізнавання ключових слів.

Таким чином, вирази (7) - (10) складають аналітичне забезпечення концептуальної моделі в галузі визначення ефективності процесу розпізнавання ключових слів у голосовому сигналі. Вони деталізують інтегральну ефективність системи та її залежність від ефективності розробки моделей, їх застосування та обробки голосових даних. Кожна зі складових ефективності, у свою чергу, розкладається на конкретні операції, такі як формування навчальних даних, вибір ознак, проектування та налаштування моделей.

Висновки та перспективи подальшого дослідження.

У даній роботі було розроблено концептуальну модель підходу до розпізнавання ключових слів у голосовому сигналі. Ця модель забезпечує детальний та структурований опис процесу розробки засобів розпізнавання ключових слів, включаючи основні компоненти, їхню взаємодію та критичні фактори ефективності.

Запропонована концептуальна модель враховує низку викликів, таких як варіативність голосових сигналів, стійкість до шумів, обмеженість навчальних даних та вимоги до продуктивності в реальному часі. Модель інтегрує сучасні методи глибокого навчання, зокрема нейронні мережі, приховані марковські моделі та методи аугментації даних для підвищення точності розпізнавання.

Окремим внеском є розробка формалізованого підходу до оцінювання ефективності системи з використанням комплексних метрик, які охоплюють точність, швидкість реагування, надійність та масштабованість. Це дозволяє всебічно аналізувати та порівнювати різні методи та архітектури для задачі розпізнавання ключових слів.

Результати даного дослідження формують основу для подальшого вдосконалення систем розпізнавання мовлення, інтеграції з додатковими моделями, такими як розпізнавання мовця, а також адаптації до специфічних областей застосування.

Перспективні напрямки дослідження включають розвиток методів трансферного навчання, покращення якості та різноманітності навчальних даних, а також оптимізацію архітектур нейронних мереж для підвищення надійності та ефективності в динамічних умовах. Особливу увагу варто

звернути на застосування та вдосконалення методів динамічного програмування, зокрема алгоритму динамічного трансформування часу (DTW), що дозволяє ефективно справлятися з варіативністю та нестабільністю голосових сигналів. Ці методи забезпечують оптимальне вирівнювання часових послідовностей, що є критично важливим для точного розпізнавання мови в реальному часі. Розширення можливостей DTW та інших алгоритмів динамічного програмування сприятиме подоланню обмежень існуючих систем та підвищенню їхньої адаптивності до складних умов експлуатації.

Список бібліографічного опису

1. Umesh Dwivedia, T., Guptab, S., Upadhyayb, S. K., Shuklab, Y., & Ahujab, S. Automatic Speech Recognition System Using Hybrid Hidden Markov Model and Human Emotion Recognition System.
2. Rashmi, S., Hanumanthappa, M., & Reddy, M. V. (2018). Hidden Markov Model for speech recognition system—a pilot study and a naive approach for speech-to-text model. In Speech and Language Processing for Human-Machine Communications: Proceedings of CSI 2015 (pp. 77-90). Springer Singapore.
3. Gunawan, A. (2010). English digits speech recognition system based on hidden Markov models. In Proceedings of International Conference Computer.
4. Deshmukh, A. M. (2020). Comparison of hidden markov model and recurrent neural network in automatic speech recognition. European Journal of Engineering and Technology Research, 5(8), 958-965.
5. Khurana, S., Laurent, A., Hsu, W. N., Chorowski, J., Lancucki, A., Marxer, R., & Glass, J. (2020). A convolutional deep markov model for unsupervised speech representation learning. arXiv preprint arXiv:2006.02547.
6. Chen, K. Y., Tsai, C. P., Liu, D. R., Lee, H. Y., & Lee, L. S. (2019). Completely unsupervised speech recognition by a generative adversarial network harmonized with iteratively refined hidden markov models. arXiv preprint arXiv:1904.04100.
7. Abdalla, M. I., & Ali, H. S. (2010). Wavelet-based mel-frequency cepstral coefficients for speaker identification using hidden markov models. arXiv preprint arXiv:1003.5627.
8. Belinkov, Y., & Glass, J. (2017). Analyzing hidden representations in end-to-end automatic speech recognition systems. Advances in Neural Information Processing Systems, 30.
9. Momeni, L., Afouras, T., Stafylakis, T., Albanie, S., & Zisserman, A. (2020). Seeing wake words: Audio-visual keyword spotting. arXiv preprint arXiv:2009.01225.
10. Berg, A., O'Connor, M., & Cruz, M. T. (2021). Keyword transformer: A self-attention model for keyword spotting. arXiv preprint arXiv:2104.00769.
11. Chen, G., Parada, C., & Sainath, T. N. (2015, April). Query-by-example keyword spotting using long short-term memory networks. In 2015 IEEE international conference on acoustics, speech and signal processing (ICASSP) (pp. 5236-5240). IEEE.
12. Tereikovskyi, I., Hu, Z., Chernyshev, D., Tereikovska, L., Korystin, O., & Tereikovskyi, O. (2022). The method of semantic image segmentation using neural networks. International Journal of Image, Graphics and Signal Processing, 13(6), 1.
13. Toliupa, S., Tereikovskyi, I., Tereikovskyi, O., Tereikovska, L., Nakonechnyi, V., & Kulakov, Y. (2020, February). Keyboard dynamic analysis by Alexnet type neural network. In 2020 IEEE 15th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET) (pp. 416-420). IEEE.
14. Seilova, N., Tereykovskaya, L., & Nadgi, A. (2016). Conceptual model to ensure the efficiency of neural network recognition of phonemes in distance learning. Seilova N, Tereykovskaya L, Nadgi A. Vestnik KazNRTU, 114(2), 345-351.
15. Дичка, І., Терейковський, І., Самофалов, А., Терейковська, Л., & Романкевич, В. (2023). МНОЖИНА КРИТЕРІЇВ ЕФЕКТИВНОСТІ ФОРМУВАННЯ БАЗ ДАНИХ ЕМОЦІЙНО ЗАБАРВЛЕНИХ ГОЛОСОВИХ СИГНАЛІВ. Електронне фахове наукове видання «Кібербезпека: освіта, наука, техніка», 1(21), 65-74.
16. I. A. Dychka, I. A. Tereikovskyi, O. S. Korovii, L. O. Tereikovska, and V. O. Romankevych, "Evaluation of the effectiveness of means for recognizing the emotional tonality of text fragments," Scientific notes of Taurida National V.I. Vernadsky University. Series: Technical Sciences, vol. 34 (73), no. 3, part 1, pp. 130-135, 2023.
17. Дичка, І.А., Терейковський, І.А., Дідус, А.В., Терейковська, Л.О., & Бояринова, Ю.Є. (2023). Оцінка ефективності засобів розпізнавання ключових слів у голосовому сигналі. Вчені записки. https://doi.org/10.32782/2663-5941/2023.3.1/19

References

1. Umesh Dwivedia, T., Guptab, S., Upadhyayb, S. K., Shuklab, Y., & Ahujab, S. Automatic Speech Recognition System Using Hybrid Hidden Markov Model and Human Emotion Recognition System.
2. Rashmi, S., Hanumanthappa, M., & Reddy, M. V. (2018). Hidden Markov Model for speech recognition system—a pilot study and a naive approach for speech-to-text model. In Speech and Language Processing for Human-Machine Communications: Proceedings of CSI 2015 (pp. 77-90). Springer Singapore.
3. Gunawan, A. (2010). English digits speech recognition system based on hidden Markov models. In Proceedings of International Conference Computer.
4. Deshmukh, A. M. (2020). Comparison of hidden markov model and recurrent neural network in automatic speech recognition. European Journal of Engineering and Technology Research, 5(8), 958-965.
5. Khurana, S., Laurent, A., Hsu, W. N., Chorowski, J., Lancucki, A., Marxer, R., & Glass, J. (2020). A convolutional deep markov model for unsupervised speech representation learning. arXiv preprint arXiv:2006.02547.

6. Chen, K. Y., Tsai, C. P., Liu, D. R., Lee, H. Y., & Lee, L. S. (2019). Completely unsupervised speech recognition by a generative adversarial network harmonized with iteratively refined hidden markov models. arXiv preprint arXiv:1904.04100.
7. Abdalla, M. I., & Ali, H. S. (2010). Wavelet-based mel-frequency cepstral coefficients for speaker identification using hidden markov models. arXiv preprint arXiv:1003.5627.
8. Belinkov, Y., & Glass, J. (2017). Analyzing hidden representations in end-to-end automatic speech recognition systems. Advances in Neural Information Processing Systems, 30.
9. Momeni, L., Afouras, T., Stafylakis, T., Albanie, S., & Zisserman, A. (2020). Seeing wake words: Audio-visual keyword spotting. arXiv preprint arXiv:2009.01225.
10. Berg, A., O'Connor, M., & Cruz, M. T. (2021). Keyword transformer: A self-attention model for keyword spotting. arXiv preprint arXiv:2104.00769.
11. Chen, G., Parada, C., & Sainath, T. N. (2015, April). Query-by-example keyword spotting using long short-term memory networks. In 2015 IEEE international conference on acoustics, speech and signal processing (ICASSP) (pp. 5236-5240). IEEE.
12. Tereikovskyi, I., Hu, Z., Chernyshev, D., Tereikovska, L., Korystin, O., & Tereikovskyi, O. (2022). The method of semantic image segmentation using neural networks. International Journal of Image, Graphics and Signal Processing, 13(6), 1.
13. Toliupa, S., Tereikovskyi, I., Tereikovskyi, O., Tereikovska, L., Nakonechnyi, V., & Kulakov, Y. (2020, February). Keyboard dynamic analysis by Alexnet type neural network. In 2020 IEEE 15th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET) (pp. 416-420). IEEE.
14. Seilova, N., Tereykovskaya, L., & Nadgi, A. (2016). Conceptual model to ensure the efficiency of neural network recognition of phonemes in distance learning. Seilova N, Tereykovskaya L, Nadgi A. Vestnik KazNRTU, 114(2), 345-351.
15. Dychka, I., Tereikovskyi, I., Samofalov, A., Tereikovska, L., & Romankievych, V. (2023). A Set of Effectiveness Criteria for Creating Databases of Emotionally Colored Voice Signals. Electronic Professional Scientific Publication "Cybersecurity: Education, Science, Technique", 1(21), 65-74.
16. I. A. Dychka, I. A. Tereikovskyi, O. S. Korovii, L. O. Tereikovska, and V. O. Romankevych, "Evaluation of the effectiveness of means for recognizing the emotional tonality of text fragments," Scientific notes of Taurida National V.I. Vernadsky University. Series: Technical Sciences, vol. 34 (73), no. 3, part 1, pp. 130-135, 2023.
17. Dychka, I. A., Tereikovskyi, I. A., Didus, A. V., Tereikovska, L. O., & Boyarynova, Y. Ye. (2023). Evaluation of the Effectiveness of Tools for Recognizing Keywords in Voice Signals. Scholarly Notes. https://doi.org/10.32782/2663-5941/2023.3.1/19