ВИКОРИСТАННЯ ШТУЧНОГО ІНТЕЛЕКТУ В СИСТЕМАХ ГОЛОСОВОЇ АВТЕНТИФІКАЦІЇ

Пастушенко М.С., Петраченко М.О.

Кафедра інфокомунікаційної інженерії ім. В.В. Поповського,
Харківський національний університет радіоелектроніки,
Україна

E-mail: mykola.pastushenko@nure.ua,
maksym.petrachenko@nure.ua

Abstract

Voice authentication is becoming more and more relevant in the modern world as a convenient and reliable method of personal identification. Artificial intelligence and phase modulation are promising technologies for improving this system. The purpose of this paper is to study the use of artificial intelligence and phase modulation in voice authentication systems to improve the quality and reliability of the face identification process. To perform this study, analytical work was conducted, including literature analysis and a review of existing voice authentication systems. Additionally, machine learning methods, discrete Fourier transform, and integration of phase information into neural networks were investigated.

Машинне навчання є ключовим методом у класифікації голосових сигналів для розпізнавання мови, що сприяє застосуванням у віртуальних асистентах, обслуговуванні клієнтів та транскрипції. Воно включає в себе алгоритми, які розвиваються та покращують свою продуктивність при обробці більшої кількості даних, ідеально підходячи для завдань розпізнавання мови [1].

До класифікації та ідентифікації голосових сигналів відносяться такі техніки машинного навчання:

1) Навчання з вчителем використовує позначені дані для навчання моделей для завдань, таких як перетворення мови в текст і ідентифікація розмовника.

2) Навчання без вчителя ідентифікує патерни в непозначених даних, корисно для групування схожих голосових сигналів.

3) Неповне навчання поєднує позначені та непозначені дані, забезпечуючи баланс між точністю та зусиллями з позначенням даних.

4) Підсилення навчання оптимізує продуктивність моделі з часом через систему, яка базується на винагороді.

Алгоритми машинного навчання, які використовуються в розпізнаванні голосу, включають:

5) Моделі глибоких нейронних мереж (Deep Neural Networks, DNN) моделюють складні патерни голосових даних.

6) Згорткові нейронні мережі (Convolutional Neural Network, CNN) ефективно обробляють спектральні патерни голосу, ідентифікуючи локальні особливості.

7) Рекурентні нейронні мережі (Deep Neural Networks, RNN) відмінно розуміють сказану мову, оброблюючи вхідні дані послідовно.

8) Нейронні мережі з довгостроковою пам'яттю (Long Short-Term Memory Networks, LSTM) подолують проблеми зі зникненням градієнту, покращуючи продуктивність.

9) Воротні рекурентні блоки (Gated Recurrent Unit, GRU) пропонують спрощені та швидкі альтернативи LSTM, завдяки чому підходять для різних завдань.

10) Машини опорних векторів (Support Vector Machine, SVM) класифікують голосові сигнали на основі виділених ознак, особливо корисні в умовах високо-вимірних даних.

11) K-найближчих сусідів (K-Nearest Neighbors, K-NN) є простим алгоритмом для класифікації та регресії, який ґрунтується на метриках подібності.

12) Класифікатори наївного Байєса застосовують теорему Байєса та припущення про незалежність ознак для класифікації голосових сигналів.

Видобування ознак та попередня обробка є невід'ємною частиною [2], але вибір методів та алгоритмів машинного навчання значущо впливає на продуктивність системи розпізнавання голосу. Глибоке розуміння цих технік є важливим для ефективної розробки систем розпізнавання голосу.

Для цифрової обробки сигналів застосовується дискретне перетворення Фур'є (Discrete Fourier Transform, DFT) до періодично дискретизованої часової області [3]. Тут будь-який цифровий звук f (сигнал, який може бути розкладений на різні компоненти частот) є сумою чистих звуків на різних частотах. Застосування DFT до неперервних сигналів включає кілька кроків, як показано на рисунку 1.

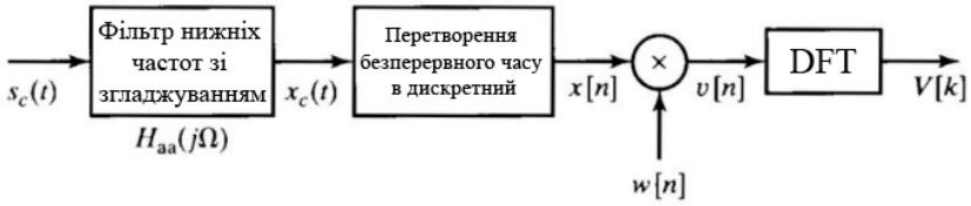

Рис. 1. Етапи обробки в дискретному Фур'є-аналізі неперервного сигналу в часі

Для аналізу та обробки звукових сигналів застосовуються методи, які використовують перетворення Фур'є [4]. На першому етапі застосовується згладжувальний фільтр, щоб підготувати сигнал. Далі, сигнал розбивається на вікна і згладжується. Після цього обчислюється DFT для отримання спектру сигналу. Відредагований сигнал може бути використаний для видалення шуму або небажаних компонентів, поліпшення якості звуку та виявлення сигналів в шумі.

Фур'є-аналіз також можна використовувати для покращення якості звуку, видаляючи небажані ноти або шум. Наприклад, дратівливий високо-частотний звук у цифровому звукозаписі можна ідентифікувати за стрибком у перетворенні Фур'є звукової хвилі. Зменшивши цей пік і застосувавши зворотне перетворення Фур'є, можна отримати відредагований цифровий звук, вільний від небажаних нот.

У сигнальній обробці фаза відноситься до часової відповідності складових синусоїдального сигналу відносно посилання, як правило, початку хвилі. Фаза впливає на сприйняття звуку, впливаючи на підсилення або приглушення звуку, коли складові знаходяться в або поза фазою, що призводить до втручання у звукову хвилю.

Фаза також впливає на тембр, оскільки складові, що знаходяться на 180 градусів [5] відносно одне одного, створюють різні тембральні властивості. Дані про фазу можуть бути виражені через фазовий кут або комплексну фазу, яка представляє як амплітуду, так і фазовий кут.

Останні досягнення в глибокому навчанні підкреслюють важливість фазових даних в додатках, пов'язаних із голосом. Фаза вкладає в себе часову структуру звуку, яка є важливою для завдань, таких як розпізнавання мови та аналіз музики. Складні представлення аудіосигналів, які комбінують амплітуду і фазу, ідеально підходять для різних обробкових завдань, як це підтверджують складні спектрограми.

Фазові дані покращують якість мови, збільшуючи чіткість в шумних середовищах. Вони сприяють розділенню музичних та звукових джерел, як у музичному реміксі. Надійне визнання розмовника великою мірою користується фазовими даними, зменшуючи чутливість до шуму та відлучення. У синтезі мови та конвертації голосу збережені фазові дані дозволяють плавні переходи між кадрами, підвищуючи природність та чіткість.

Стратегії інтеграції фазових даних в нейронні мережі включають в себе використання складних спектрограм як вхідних даних [6], відновлення фази, втрати, зв'язані з фазою, та спільні моделі амплітуди і фази. Проблема фазової неоднозначності, коли декілька конфігурацій фази дають однаковий амплітудний спектр, залишається предметом досліджень з потенціалом для подальших вдосконалень.

Ефективні архітектури нейронних мереж для обробки фази в реальному часі та загального застосування в різних акустичних умовах та голосових завданнях є актуальними напрямками дослі-

джень. Інтеграція фазових даних в нейронні мережі перевершила обробку голосу, із застосуванням від покращення якості мови до розділення звукових джерел.

Фазова модуляція є ключовою складовою голосового сигналу, яка раніше ігнорувалася при аналізі та обробці голосових даних. Використання фазової інформації в голосовій автентифікації дозволяє отримати кілька переваг [7].

Кепстральні характеристики, такі як Мел-частотні кепстральні коефіцієнти (Mel Frequency Cepstral Coefficients, MFCC), зазвичай використовуються для аналізу амплітудно- частотних характеристик голосового сигналу. Включення фазової інформації дозволяє покращити якість цих характеристик. Фазова модуляція може допомогти врахувати динаміку голосу, а не тільки його частотно-амплітудні особливості.

Коефіцієнти використовуються для моделювання формантації голосу. Використання фазової інформації може поліпшити точність визначення цих коефіцієнтів, що впливає на точність голосової автентифікації.

Використання фазової модуляції може дати результати нічим не гірше штучного інтелекту. Це важливий аспект, оскільки штучний інтелект також використовується для голосової автентифікації. Фазова модуляція дозволяє отримати більше інформації з голосового сигналу, що може покращити точність ідентифікації особи. Вона може бути більш стійкою до змін у голосовому сигналі, таким як зміни в акценті, настрої або шумі.

Використання фазової інформації може допомогти уникнути проблем, пов'язаних з обробкою розбіжностей, які можуть виникнути при використанні тільки частотно-амплітудних характеристик. Незважаючи на важливість фазової модуляції, штучний інтелект також залишається потужним інструментом для голосової автентифікації. Об'єднуючи обидва підходи, можливо досягти ще кращої точності та надійності системи автентифікації. Використання фазової модуляції в голосовій автентифікації може дійсно покращити результати і надати конкурентну альтернативу традиційній обробці частотно-амплітудних характеристик.

*Л*ітература

1. Karpagavalli R., Balamurugan K. Research on voice authentication systems: technologies, challenges and opportunities. International Journal of Computer Science and Mobile Computing. 2016. 48 p.

2. Qian K., Zhang Z., Ren J., Han Z. Robust opinion enhancement based on deep learning for voice authentication. IEEE Transactions on Information Forensics and Security. 2021. 32 p.

3. Norton B., Louridas P., Spinellis D., Kakarontzas G. The future of voice interfaces. Journal of Systems and Software. 2022. 15 p.

4. Smith J. Voice Authentication using Amplitude-Frequency Information. In: Proceedings of the International Conference on Biometrics, 2018. P. 45 – 52.

5. Johnson, M. Advances in Voice Recognition Technology. Journal of Acoustic Signal Processing, 2022. P. 112 – 125.

6. Smith, J. R., & Davis, A. L. (2019). Leveraging Phase Information in Deep Learning Models for Robust Speaker Recognition. Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2019. P. 1234 – 1238.

7. M.Pastushenko, Ya.Krasnozheniuk, M. Zaika. Investigation of Informativeness and Stability of Mel-Frequency Cepstral Coefficients Estimates based on Voice Signal Phase Data of Authentication System User. International Conference Problems of Infocommunications. Science and Technology (PIC S&T′2020), 2020. P. 1 – 5.