УДК 004.522

К. Р. Салієва
О. В. Салієва

ПОРІВНЯЛЬНИЙ АНАЛІЗ МЕТОДІВ РЕАЛІЗАЦІЇ ГОЛОСОВОЇ БІОМЕТРІЇ

Вінницький національний технічний університет

Анотація

У доповіді розглянуто сучасні методи реалізації голосової біометрії, проаналізовано їхні переваги та недоліки, створено порівняльну таблицю даних методів, що ґрунтується на визначенні рангів основних характеристик.

Ключові слова: біометрія, голосова біометрія, розпізнавання голосу, метод опорних векторів, модель суміші Гауса, прихована марківська модель, динамічне викривлення часу, векторне квантування, штучні нейронні мережі.

Annotation

The report considers modern methods of the implementation of voice biometrics, analyzes their advantages and disadvantages, created a comparative table of these methods, based on determining the ranks of the main characteristics.

Keywords: biometrics, voice biometrics, voice recognition, support vector machine, Gaussian mixture model, hidden markov model, dynamic time warping, vector quantization, artificial neural networks.

Вступ

Біометрія визначається як унікальна, вимірювальна, біологічна характеристика або ознака певного індивіда, яка автоматично розпізнає або підтверджує особистість людини. Наука про біометрію полягає в статистичних аналізах біометричних характеристик, які наразі широко використовуються в різних сферах діяльності. На сьогодні існує п’ять найпоширеніших біометричних моделей, що базуються на аутентифікації за: відбитком пальця, геометрією руки, голосом, райдужною оболонкою чи сітківкою ока, геометрією чи термограмою обличчя.

Біометрія голосу – метод, який включає ідентифікацію та перевірку за допомогою обробки голосового сигналу. Цей метод має унікальну перевагу над іншими біометричними методами, оскільки він базується на мовленні – основному векторі комунікації, і особливо важливий у телефонних діалогових системах, де він є цілком природним. За останні роки голосова біометрія стала все більш популярною в різноманітних додатках, таких як Google та Siri. Крім того, вона використовується для удосконалення захисту конфіденційної інформації в різноманітних сферах людської діяльності, зокрема: судовій експертизі, фінансах, телекомунікаціях та інших [1, 2].

У міру розвитку аутентифікації на основі голосу з’явилися два різних підходи [2]:

– перший полягає в тому, щоб людина повторила одне й те ж речення декілька разів та створила аналогічний шаблон, що складається з діапазону голосових відбитків. Недолік цього підходу в тому, що голосовий друк є більш загальним;

– другий підхід – створення голосового відбитку за допомогою однієї фрази або слова і збереження лише цього окремого голосового відбитка. Недоліком є те, що третя особа може записати спробу аутентифікації та відтворити її, щоб отримати доступ.

Технологія розпізнавання динаміків, яка аналізує та моделює відбитки голосу, була важливою дослідницькою роботою протягом останніх десятиліть, але є ще багато проблем, які потрібно вирішити. Важливою проблемою в багатьох областях обробки мовлення є визначення наявності мовних періодів у даному сигналі. Цю задачу можна представити, як проблему гіпотези, метою якої є визначення, категорії чи класу до якої належить даний сигнал. При створенні голосової автентифікації необхідно проаналізувати наявні математичні методи, що допоможуть покращити розпізнавання голосу мовця.

Метою даної роботи є порівняльний аналіз відомих математичних методів для реалізації голосової біометрії.

Результати дослідження

Наразі існує два основних етапи для розпізнавання мовних шаблонів: навчання та порівняння. Для цього необхідно добре сформувати математичні основи та створити послідовність представлення шаблонів для його надійного порівняння. Існують такі підходи до реалізації голосової біометрії [3]:

– на основі шаблону – при авторизації особи її мовлення порівнюється з набором попередніх записів для знаходження найкращої відповідності. Перевагою є те, що даний підхід досить продуктивний для пошуку точних моделей, а недоліком – фіксування раніше записаних шаблонів;

– стохастичний підхід – відхилення в мовленні моделюється статистично. Перевага його в тому, що використовується автоматична процедура статистичного навчання. Недоліком є те, що алгоритм повинен приймати попередні припущення в моделюванні, що може призвести до помилкового рішення, негативно впливаючи на продуктивність системи;

– підхід на основі знань (штучний інтелект) – поєднання акустичного фонетичного підходу та розпізнавання образів. Даний підхід варто застосувати для моделювання варіацій у мовленні, але такі знання важко здобути та уміло використати для успішної реалізації, тому цей підхід не є практичним.

Під час дослідження математичних алгоритмів було вирішено виокремити та розглянути шість методів, що найчастіше використовуються для реалізації голосової біометрії.

1. Метод опорних векторів (Support Vector Machine (SVM)) – це набір методів навчання з наглядом, що використовуються для класифікації, регресії та виявлення викидів. SVM є одним із найпопулярніших алгоритмів контрольованого навчання, який використовується для задач класифікації та регресії. Однак, насамперед, він використовується для задач класифікації в машинному навчанні. Метою алгоритму SVM є створення найкращої лінії або межі рішення, яка може розділити n-вимірний простір на класи, що сприятиме легкому розміщенню нової точки даних у правильну категорію. Ця межа найкращого рішення називається гіперплощиною. Основні переваги SVM: ефективність у великих просторах та у випадках, коли кількість вимірів переважає над кількістю зразків; використання підмножини навчальних точок у функції прийняття рішень; різні функції ядра можна вказати для функції прийняття рішень; надаються звичайні ядра і можна вказати власні. До недоліків можна віднести те, що коли кількість функцій набагато більше, ніж кількість зразків, необхідно уникати підбору функцій ядра, і термін її регуляризації є вирішальним; SVM не надають безпосередньої оцінки ймовірності, вони виконуються за допомогою дорогої п’ятикратної перехресної перевірки [4].

2. Модель суміші Гаусса (Gaussian Mixture Model (GMM)) – це імовірнісна модель, яка передбачає, що всі точки даних генеруються на основі суміші гауссових розподілів з невідомими параметрами. Модель гауссової суміші може бути використана для: кластеризації, яка є завданням групування набору точок даних у кластери; пошуку кластерів у наборах даних, де кластери не можуть бути чітко визначені; оцінювання ймовірності того, що нова точка даних належить кожному кластеру. Гауссові моделі суміші також відносно стійкі до викидів, це означає, що вони все ще можуть давати точні результати, навіть якщо є деякі точки даних, які однозначно не вписуються в жоден із кластерів. Це робить GMM гнучким і потужним інструментом для кластеризації даних. Переваги моделі: найшвидший алгоритм для вивчення моделей сумішей; алгоритм не зміщує середнє значення до нуля або не зміщуватиме розміри кластерів, тому що він максимізує лише ймовірність. До недоліків можна віднести те, що при недостачі точок, оцінювання матриць стає важким процесом, алгоритм розходиться і знаходить рішення з нескінченною ймовірністю; цей алгоритм буде завжди використовувати всі компоненти, до яких має доступ, потребуючи закритих даних або теоретичних критеріїв інформації [5].

3. Прихована марківська модель (Hidden Markov Model (HMM)) – це модель, в якій спостерігається послідовність викидів, але невідомою є послідовність станів, через які пройшла модель, щоб створити викиди. Аналіз прихованих марківських моделей спрямований на відновлення послідовності станів зі спостережуваних даних. При правильному застосуванні цієї моделі для вирішення ряду важливих питань та прикладних задач може привести до позитивних результатів. Для HMM є такі проблеми: враховуючи параметри моделі та спостережувані дані, необхідно оцінити оптимальну послідовність прихованих станів і розрахувати ймовірність даних; враховуючи лише спостережувані дані, потрібно оцінити параметри моделі. Першу та другу задачу можна вирішити за допомогою алгоритмів динамічного програмування,

відомих як алгоритм Вітербі та алгоритм «Вперед-Назад», відповідно. Останній може бути вирішений за допомогою ітераційного алгоритму максимізації очікування (EM), відомого як алгоритм Баума-Велча [6].

4. Метод динамічного викривлення часу (Dynamic Time Warping (DTW)) – є одним з алгоритмів для вимірювання подібності двох тимчасових рядів, які можуть відрізнятися за швидкістю. Метою методів порівняння часових рядів є отримання метрики відстані між двома вхідними часовими рядами. Подібність або несхожість подвійних рядів, зазвичай, обчислюється шляхом перетворення даних у вектори та обчислення евклідової відстані між цими точками у векторному просторі. DTW дає нелінійне (пружне) вирівнювання між подвійними рядами. Даний метод шукає найкраще узгодження між дворазовими рядами. Це створює більш інтуїтивну міру подібності, що дозволяє схожим фігурам збігатися, навіть якщо вони не зберігаються за фазою на осі часу. Добре використовується для поєднання зразка голосової команди з командою інших, навіть якщо людина говорить швидше або повільніше, ніж попередньо записаний зразок голосу [7].

5. Векторне квантування (Vector Quantization (VQ)) – це блочний метод просторової області, який став дуже популярним з початку 1980-х років. У VQ вхідні дані зображення спочатку розкладено на k-вимірні вхідні вектори. Дуже важливою проблемою у VQ є дизайн кодової книги. Для розв’язання цієї проблеми найчастіше використовується алгоритм Лінде-Бузо-Грея (LBG), який є узагальненням алгоритму Ллойда-Макса для скалярного квантування. Векторне квантування має продуктивність, яка конкурує з продуктивністю перетворення кодування. Хоча складність декодера є незначною (таблиця пошуку), висока складність кодера та високі вимоги до пам’яті методу все ще обмежують його використання на практиці. Як і кодування трансформації, VQ має проблему блокування артефактів на дуже низьких швидкостях [8].

6. Штучні нейронні мережі (Artificial Neural Networks (ANN)) – це обчислювальна модель, що складається з кількох елементів обробки, які отримують вхідні дані та видають вихідні дані на основі їх попередньо визначених функцій активації. Ці мережі імітують біологічну нейронну мережу, але використовують скорочений набір концепцій біологічних нейронних систем. Зокрема, моделі ANN моделюють електричну активність мозку та нервової системи. Елементи обробки (також відомі як нейрод або персептрон) з’єднані з іншими елементами обробки. Зазвичай, нейроди розташовуються в шарі або векторі, при цьому вихід одного шару служить вхідним сигналом для наступного шару і, можливо, інших шарів. Нейрод може бути з'єднаний з усіма або підгрупою нейродів у наступному шарі, при цьому ці з'єднання імітують синаптичні зв'язки мозку. Сигнали зважених даних, що надходять у нейрод, імітують електричне збудження нервової клітини та передачу інформації всередині мережі або мозку. Теоретично, для моделювання асинхронної діяльності нервової системи людини елементи обробки штучної нейронної мережі також повинні бути активовані зваженим вхідним сигналом асинхронно. Однак більшість програмних і апаратних реалізацій штучних нейронних мереж реалізують більш дискретизований підхід, який гарантує, що кожен елемент обробки активується один раз для кожного представлення вектора вхідних значень [9].

Оцінювання методів у роботі проводилося за методом ранжування за шкалою важливості показників (від 1 до 10, чим вище значення показника, тим краще). Результати наведено в таблиці 1.

Таблиця 1 – Порівняльний аналіз методів голосової біометрії

Показники характеристик	SVM	GMM	HMM	DTW	VQ	ANN
Відмова у доступі справжньому користувачу	7	8	6	9	5	8
Помилкова ідентифікація «чужого голосу»	8	7	5	8	6	8
Стійкість до підробок голосу та атак	7	7	7	8	5	9
Похибка при розпізнаванні	9	7	6	9	7	9
Розпізнання голосу при хворобах користувача	7	6	5	8	5	8
Час розпізнавання користувача	8	7	7	9	6	9
Розмір голосового шаблону	7	6	5	8	3	7
Простота реалізації методу	8	6	5	7	4	3
Вартість реалізації	9	6	5	8	7	4

Висновки

В останні роки зростає інтерес до використання біометричних характеристик, як засобу розпізнавання та ідентифікації особи. Людський голос є одним з найважливіших біометричних ідентифікаторів людини.

У результаті порівняльного аналізу біометричних методів було встановлено, що найбільш перспективним для подальшого дослідження виявився метод опорних векторів для створення біометричної ідентифікації. Даний метод вирізняється своєю дешевизною і простотою реалізації, тому що не потребує додаткового дорогого устаткування. Також він дозволяє контролювати доступ до конфіденційної інформації.

У зв'язку із актуальністю загрози запису та відтворення голосового відбитку варто поєднати даний метод із MFCC для виділення коефіцієнту та VAD для визначення сегменту сигналу. В майбутніх дослідженнях можна поєднати метод SVM і з іншими методами, що наведені в даній доповіді. Сьогодні найчастіше використовують поєднання SVM/HMM та SVM/GMM. Це значною мірою покращить та удосконалить алгоритм голосової біометрії, що буде ефективним при використанні звичайним користувачем та зменшить вірогідність зламу третьою особою.

Отже, в подальших дослідженнях доцільно сконцентрувати увагу на підвищенні якості ідентифікації за допомогою методу SVM для створення голосової біометрії.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Bio-metric Encryption of Data Using Voice Recognition [Електронний ресурс] // Science Publishing Group. – 2021. – Режим доступу до ресурсу: http://www.sciencepublishinggroup.com/j/acis.

2. An Overview and Analysis of Voice Authentication Methods [Електронний ресурс]. – 2016. – Режим доступу до ресурсу: https://www.semanticscholar.org/paper/An-Overview-and-Analysis-of-Voice-Authentication-Shoup-Talkar/572af444f0382b8e7e156ab36192da95a3b8dec4.

3. Efficient voice activity detection algorithms using long-term speech information [Електронний ресурс] // Speech Communication. – 2014. – Режим доступу до ресурсу: https://www.journals.elsevier.com/speech-communication.

4. Support Vector Machines [Електронний ресурс]. – 2020. – Режим доступу до ресурсу: https://scikit-learn.org/stable/modules/svm.html.

5. Gaussian Mixture Models: What are they & when to use? [Електронний ресурс]. – 2022. – Режим доступу до ресурсу: https://vitalflux.com/gaussian-mixture-models-what-are-they-when-to-use/.

6. Hidden Markov Models (HMM) [Електронний ресурс]. – 2022. – Режим доступу до ресурсу: https://www.mathworks.com/help/stats/hidden-markov-models-hmm.html.

7. Dynamic Time Warping (DTW) [Електронний ресурс]. – 2019. – Режим доступу до ресурсу: https://medium.datadriveninvestor.com/dynamic-time-warping-dtw-d51d1a1e4afc.

8. Vector Quantization [Електронний ресурс]. – 2003. – Режим доступу до ресурсу: https://www.sciencedirect.com/topics/engineering/vector-quantization.

9. Artificial Neural Network [Електронний ресурс] // Journal of Environmental Management. – 2015. – Режим доступу до ресурсу: https://www.sciencedirect.com/topics/earth-and-planetary-sciences/artificial-neural-network.

Салієва Катерина Рустамівна - студентка групи КІТС-18б, факультет менеджменту інформаційної безпеки, Вінницький національний технічний університет, м. Вінниця, e-mail: kate228778@gmail.com

Салієва Ольга Володимирівна – доктор філософії (PhD) за спеціальністю 125 «Кібербезпека», старший викладач кафедри менеджменту та безпеки інформаційних систем, Вінницький національний технічний університет, Вінниця, e-mail: salieva8257@gmail.com

Kateryna Salieva R. - student of the KITS-18b group, Faculty of Management Information Security, Vinnitsa National Technical University, Vinnitsa, email: kate228778@gmail.com

Saliieva Olha V. – Doctor of Philosophy (PhD) in 125 "Cybersecurity", Senior Lecturer, Department of Management and Security of Information Systems, Vinnytsia National Technical University, Vinnytsia, e-mail: salieva8257@gmail.com