УДК 004.056.53
DOI: 10.20535/SRIT.2308-8893.2019.4.08

ПОРІВНЯННЯ ЕФЕКТИВНОСТІ КЛАСИФІКАТОРІВ МАШИННОГО НАВЧАННЯ У КОНТЕКСТІ ГОЛОСОВОЇ БІОМЕТРІЇ

В.Я. ДАНИЛОВ, Я.В. ГРУШКО

Анотація. Порівняно сім популярних класифікаторів Python-бібліотеки scikit-learn у контексті ефективності роботи системи голосової біометрії. Для виділення векторів ознак голосу особи, що верифікується, застосовано метод MFCCs (Mel Frequency Cepstral Coefficients). У дослідженні використано такі класифікатори: K-NN (K-Nearest neighbours classifier), MLP (Multilayer perceptron), SVM (Support vector machine), DTC (Decision tree classifier), GNB (Gaussian Naive Bayes classifier), ABC (AdaBoost classifier), RFC (Random forest classifier). Як аналізовану вибірку взято голосові зразки 40 осіб тривалістю в середньому дев'ять хвилин на особу. Критерії ефективності класифікаторів вибрано відповідно до потреб систем голосової біометрії. У межах роботи виконано моделювання шахрайства у процесі аутентифікації. Найефективнішим у голосовому розпізнаванні виявився класифікатор K-NN, який за нульової кількості неправильно допущених осіб, забезпечив на 3–85% вищу точність верифікації, ніж інші класифікатори.

Ключові слова: голосова біометрія, MFCC, порівняння класифікаторів, класифікатор K-NN, машинне навчання, штучний інтелект.

ВСТУП

Натепер стало популярним поняття «штучний інтелект» як серед науковців, так і серед програмістів. Зазвичай під цим поняттям мають на увазі здатність комп'ютерів до певного сприйняття, мислення та дій, притаманних людині. Сюди можна віднести такі популярні течії штучного інтелекту, як машинний зір (CV — Computer Vision), оброблення природної мови (NLP — Natural-Language processing), голосова біометрія (VB — Voice Biometrics) та ін. Зауважимо, що NLP у поєднанні з VB можуть слугувати для підвищення комфорту комунікації між людьми з обмеженими можливостями та комп'ютером.

Яскравим прикладом актуальності досліджень з VB є опитування 500 компаній із США та Європи, проведене Pindrop у 2018 р., згідно з яким 57% з опитаних планують упровадити найближчим часом голосові технології для більшого комфорту клієнтів, а 28% уже застосовують такі технології [1].

У світі вже існує досить багато готових рішень голосової біометрії, запропонованих рядом комерційних компаній, серед яких можна виділити такі великі компанії, як Nuance, Agnitio, VoiceVault. Але, на жаль, теоретична база цих розробок у відкритий доступ майже (або зовсім) не надходить. Саме тому цю роботу спрямовано на підвищення конкурентоспроможності української наукової спільноти в галузі штучного інтелекту — Voice Biometrics.

СТАН ДОСЛІДЖЕННЯ В УКРАЇНІ ТА ЗА КОРДОНОМ

Огляд сучасної літератури свідчить, що тематика систем голосової біометрії не дуже популярна в україномовному науковому середовищі. Попри це, праці, присвячені даній тематиці, є цікавими. Так, у праці [3] В.П. Захарова та О.І. Зачека описано особливості застосування ідентифікації за голосом порівняно з іншими методами біометрії (ідентифікацією за відбитками пальців, сітківкою ока, клавіатурним почерком, венами руки, термографічною картиною обличча, ДНК тощо). Переваги мультимодальної біометрії (за голосом та обличчям особи) та огляд інших видів біометрії досліджено у дисертації [4] Ю.О. Кумченка. Однак основну увагу автор приділив саме біометрії за обличчям. Безперечно цікавою є також праця [5] О.А. Мясіщева, де описано програму на операційній системі Android, за допомогою якої можна віддалено керувати контролером Arduino через Bluetooth, подаючи відповідні голосові команди.

Дещо ближчою за проблематикою пропонованої роботи є стаття [6] Є.Ю. Щербакова, основна мета якої полягала у визначенні оптимального методу та параметрів голосового розпізнавання особи. Найбільша точність розпізнавання, якої, як повідомляється, вдалося досягнути, становить 98,6%. Однак вибірка голосових зразків тут складалась лише з 10 осіб і кількість класифікаторів дорівнювала трьом. Крім того, автор не врахував похибки другого роду — найістотнішого параметра в роботі системи у режимі захисту від несанкціонованого доступу.

Огляд зарубіжної літератури за тематикою систем голосової біометрії також не виявив великої кількості праць із вільним доступом до параметрів ідентифікаціїї та верифікації. Утім такі праці є і вони цікаві. Так, наприклад, у [7] H.N.M. Shah і M.Z. Ab Rashid побудували систему голосової біометрії, в основу якої покладено метод MFCCs отримання голосового відбитка, а також описали такі методи, як SVM (Support Vector Machine), GMM (Gaussian Mixture Model), VQ (Vector Quantization) і DTW (Dynamic Time Warping). Метод MFCC опробовано авторами на 10 голосових зразках.

У статті [8] A. Shoup, T. Talkar проаналізували наявні системи голосової аутентифікації та розробили дві власні системи. Схема першої системи включає два основні модулі: MFCCs, що був використаний для виділення векторів ознак голосу, і класифікатор GMM. Другу систему голосової біометрії виконано на базі Python-бібліотеки Dejavu [9]. На жаль, автори не надали порівняльного аналізу ефективності застосованих класифікаторів.

Отже, з огляду літератури, що є у відкритому доступі, можна зробити висновок, що питання вибору найкращого класифікатора та оптимальних параметрів ідентифікації є досить дискусійним і привертає увагу багатьох дослідників, які працюють у галузі VB.

ПОСТАНОВКА ЗАВДАННЯ

Мета роботи — порівняння семи популярних класифікаторів бібліотеки з машинного навчання scikit-learn [2] у контексті розроблення консольного додатка (мовою Python) розпізнавання людини за голосом (Voice Biometrics). Основним функціоналом такого додатка є те, що під час навчання системи голосом певної людини виділяється набір векторів голосового відбитка особи за допомогою методу MFCCs (Mel-Frequency Coefficients) [10–11], а також виконується навчання класифікатора векторами ознак. У процесі аутентифікації набір векторів ознак з голосу, отриманий за допомогою MFCCs, подається на навчену модель класифікатора для ідентифікації та верифікації особи. Обсяг вибірки голосів становить 80 зразків, які належать 40 різним дикторам. Ставиться завдання встановити найкращий класифікатор для розпізнавання.

СХЕМА РОБОТИ СИСТЕМИ

Система, що використана в дослідженні, складається з двох основних модулів: модуля MFCCs [12] отримання набору векторів ознак з голосу (теоретичну складову методу добре описав Джеймс Ліонс [13]) та модуля класифікатора бібліотеки scikit-learn [2]. Спрощену схему роботи системи в режимі навчання зображено на рис. 1.

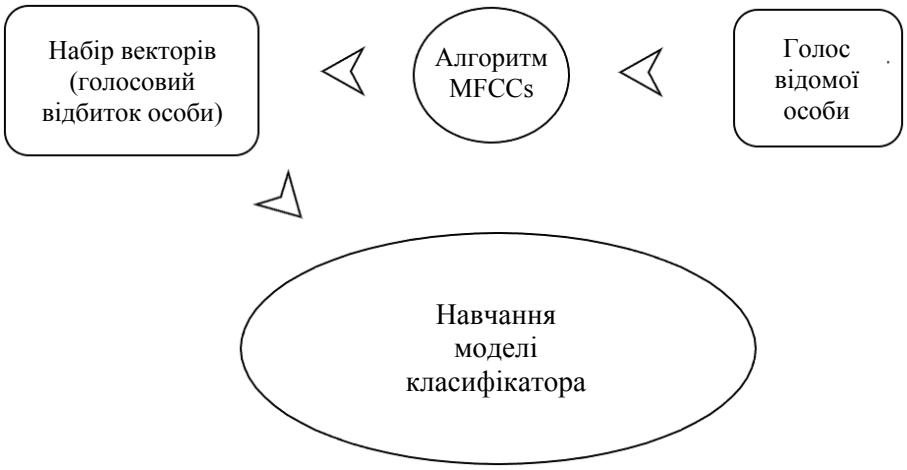

Рис. 1. Спрощена схема роботи системи в режимі навчання

Як випливає з рис. 1, навчання системи голосу особи відбувається за рахунок навчання моделі класифікатора. Спрощену схему роботи системи в режимі проходження аутентифікації невідомою особою зображено на рис. 2.

Як видно з рис. 2, ідентифікація особи виконується безпосередньо за допомогою класифікатора. Ідентифіковану особу верифікують з використанням деякого встановленого порога для певного параметра. Цей параметр ніщо інше, як відношення імовірності правильної ідентифікації найімовірнішої особи до ймовірності другої найімовірнішої особи. Для детальнішого пояснення припустімо, що система була навчена голосами п’ятьох осіб

(Ярослава, Володимира, Ігоря, Олександра, Євгенія). Тепер нехай невідома особа проходить аутентифікацію. Її голос подається на вхід системи розпізнавання, метод MFCCs виділяє з голосу вектори ознак, які подаються на класифікатор, з метою ідентифікації особи. Зрозуміло, що система зможе ідентифікувати особу, яка проходить аутентифікацію, тільки як одну з тих п'яти осіб, голосами яких вона була навчена попередньо. Нехай імовірності правильної ідентифікації невідомої особи, що відповідають відомим навченим особам, наведено у таблиці.

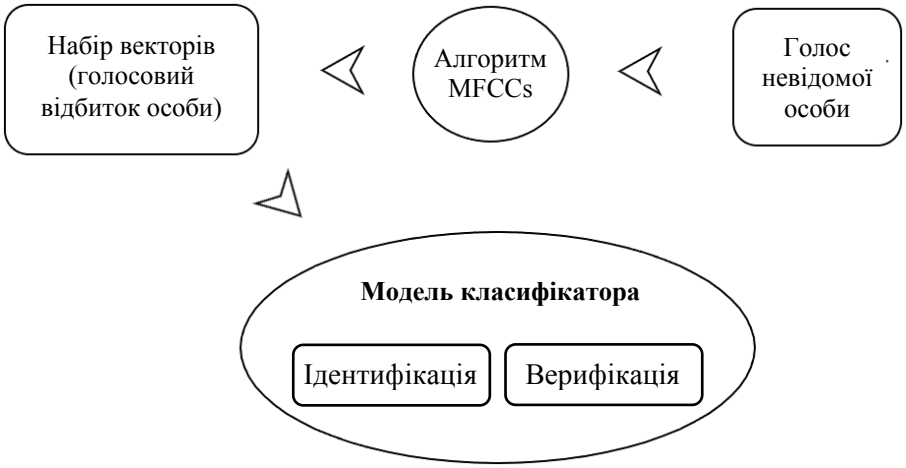

Рис. 2. Спрощена схема роботи системи в режимі аутентифікації невідомою особою

Імовірності правильної ідентифікації невідомої особи, що відповідають особам, які відомі системі

Ідентифікована особа	Імовірність правильної ідентифікації, %
Ярослав	70
Володимир	12
Ігор	8
Олександр	5
Євгеній	5

Таким чином, параметр, на базі якого відбувається верифікація, розраховується як Ident_varif = 70/12 = 5,83. Отже, чим більший цей параметр, тим більша ймовірність, що ідентифікована особа є тією самою особою, яка проходить аутентифікацію. Тому для цього параметра можна установити певний поріг (у цьому випадку не більше як 5,83). І якщо значення параметра Ident_varif більше від заданого порога, то приймається рішення, що особа верифікована правильно, тобто є тією особою, що проходить верифікацію. Інакше робиться висновок, що особа не пройшла верифікацію і не може бути допущеною до охоронних даних.

ОПИС КРИТЕРІЮ ТОЧНОСТІ КЛАСИФІКАТОРА

Відповідно до основної мети роботи порівнюється точність деяких класифікаторів, наявних у бібліотеці scikit-learn. Оскільки класифікатори розгляда-

ються як частина системи голосової біометрії, то і критерії точності мають бути актуальними для систем голосової біометрії. Для таких систем як різновидів біометрії безперечно важливою є здатність недопустити зловмисника до даних. Виміряти таку здатність можна кількістю неправильно допущених (allow-false) до закритих даних осіб, інакше кажучи, ця кількість — це кількість помилок другого роду. Зрозуміло, що чим менша кількість неправильно допущених осіб для певної вибірки, тим кращий застосований класифікатор.

Іншою, напевно, не менш важлививою здатністю систем біометрії є здатність допустити відомих системі осіб, адже система не матиме сенсу, якщо вона не буде допускати ані зловмисників, ані власників приватного контенту. Кількісно цю здатність можна виміряти кількістю правильно допущених (allow-true) осіб до закритих даних.

Маючи ці два протилежні критерії (allow-true та allow-false), можна досить точно виміряти ефективність класифікатора для верифікації. Саме ці два критерії і були використані для знаходження точності класифікаторів.

ОСОБЛИВОСТІ ВИБІРКИ

Класифікатори тестувалися на вибірці голосів, які належать 40 різним дикторам. Вибірку голосів взято з інтернет-ресурсу OpenSLR [14], де вона перебуває у вільному доступі. Як вибірку голосів використано голосові зразки 40 різних дикторів з тривалістю в середньому дев’ять хвилин. Такої тривалості виявилось цілком достатньо, щоб розбити ці зразки на менші фрагменти та утворити 160 зразків і подати їх на вхід системи, змоделювавши тим самим 80 спроб запису та аутентифікації 40 різними особами.

Особливістю цього дослідження є моделювання зловмисної атаки на систему. Моделювання реалізовано таким чином. Голосові зразки 10 осіб не подавалися на навчання системи, хоча вони подавалися на подальшу аутентифікацію, а отже, були для системи чужими. Таке моделювання щодо зловмисників є цілком виправданим, оскільки, як відомо, для будь-якої системи біометрії однією з найважливіших особливостей є здатність недопустити зловмисника до конфеденційних даних.

Таким чином, у режимі моделювання маємо лише 80 спроб виконання аутентифікації, 20 з яких напевне є зловмисними (оскільки система не була навчена зразками голосів 10 з 40 дикторів), однак система наперед не знає, які саме з цих спроб є зловмисними і повинна саме їх визначити такими. Тому максимально можлива кількість спроб правильної аутентифікації становить 60 спроб, мінімальна — 0. Максимально можлива кількість неправильної аутентифікації (кількість помилок другого роду) дорівнює 80, мінімальна — 0.

Зіставивши ці критичні кількості з конкретними показниками відповідних протестованих класифікаторів, можна зробити висновок про ефективність цих класифікаторів відповідно до завдання розпізнавання.

ПОРІВНЯННЯ СЕМИ КЛАСИФІКТОРІВ

У роботі протестовано сім популярних класифікаторів Python-бібліотеки scikit-learn [2], а саме:

1) K-NN (K-Nearest neighbours classifier – К-найближчих сусідів);
2) MLP (Multilayer perceptron – багатошаровий перцептрон);
3) SVM (Support vector machine – метод опорних векторів);
4) DTC (Decision tree classifier – класифікатор дерев ухвалення рішень);
5) GNB (Gaussian Naive Bayes classifier – наївний байєсів класифікатор);
6) ABC (AdaBoost classifier – адаптивний бустінг);
7) RFC (Random forest classifier – класифікатор випадкового лісу).

Кількісні характеристики за критеріями allow-true та allow-false цих класифікаторів зображено на рис. 3.

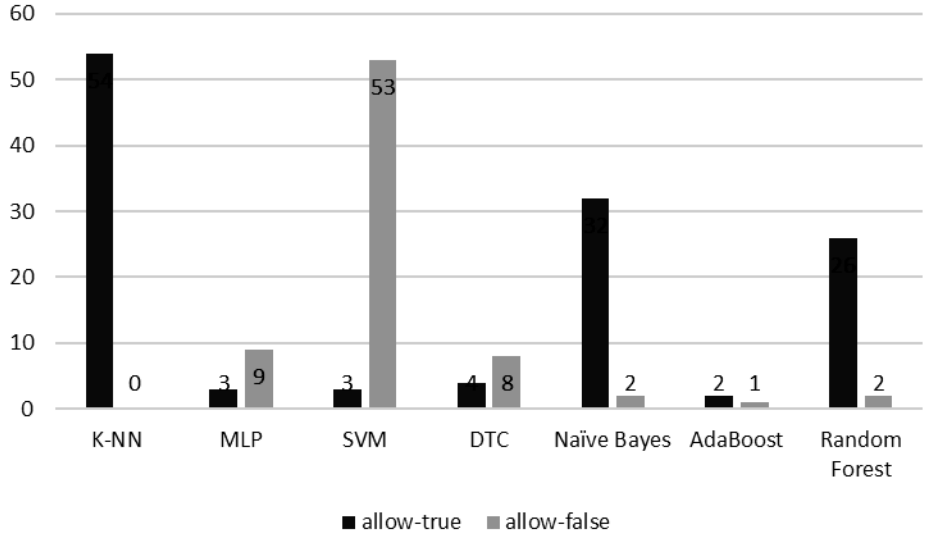

Рис. 3. Кількісні характеристики за критеріями allow-true та allow-false семи класифікаторів (у контексті додатка голосової біометрії)

Як випливає з рис. 3, найменші значення за критерієм allow-false (кількість помилок другого роду) та найбільше значення allow-true має класифікатор K-NN. Таким чином, класифікатор K-NN виявився найкращим у дослідженні. Крім того, цей класифікатор виявився точним за абсолютними показниками, набравши мінімально можливе значення кількості помилок другого роду (0 {allow-false: 0%}) і досить близьке до максимально можливого значення кількості правильно розпізнаних і верифікованих осіб (54 із 60 максимально можливих {allow-true: 90%}). Друге і третє місця посіли відповідно GNB (allow-true: 32 { на 36,7% менше за K-NN}, allow-false: 2 {на 2,5% більше за K-NN}) та RFC (allow-true: 26 {на 46,6% менше за K-NN }, allow-false: 2 {на 2,5% більше за K-NN}). Четверте місце ділять між собою одразу три класифікатори: DTC, ABC і MLP (allow-true: на 83,3% менше за K-NN, allow-false: на 10% більше за K-NN). Найгірше значення в цьому дослідженні показав метод SVM, набравши максимальне, порівняно з іншими класифікаторами, значення за критерієм allow-false (53, що на 67% більше за K-NN). Значення allow-true для клсифікатора SVM становить усього 3 (тобто на 85% менше за K-NN).

ВИСНОВКИ

Порівняно з наведеними у відкритих джерелах працях з аналізу класифікаторів, які використовуються для ідентифікації та верифікації особи у голосовій біометрії, виконане дослідження відрізняється такими характерними ознаками:

- виконано на найбільшій з відомих праць вибірці голосів осіб (40 осіб);
- застосовано моделювання шахрайства під час проходження аутентифікації;
- розраховано кількість помилок другого роду;
- проаналізовано найбільшу з відомих праць кількість класифікаторів (сім).

Описано загальну схему системи голосової біометрії, використану для тестування класифікаторів із посиланням на використані Python-бібліотеки.

У результаті дослідження встановлено, що серед семи протестованих класифікаторів python-біблотеки scikit-learn найвищі значення показників ефективності відносно як правильно допущених осіб, так і похибки другого роду має класифікатор K-NN. Абсолютні значення ефективності класифікатора K-NN (з використанням моделювання зловмисного вторгнення в систему) становлять: кількість спроб правильно допущених осіб — 54 із 60 максимально можливих (точність верифікації 90%) і кількість спроб неправильно допущених осіб — 0 (стовідсотковий захист від зловмисників).

ЛІТЕРАТУРА

1. *Pindrop 2018* voice intelligence report. — Available at: https://www.pindrop.com/2018-voice-intelligence-report/ (дата звернення: 11.11.2019).
2. *Classifier* comparison. — Available at: https://scikit-learn.org/stable/auto_examples/classification/plot_classifier_comparison.html (дата звернення: 11.11.2019).
3. *Захаров В.* Тенденції використання в діяльності правоохоронних органів біометричних технологій, які не входять до «трьох великих біометрик» / В. Захаров, О. Зачек // Наук. вісн. Львів. держ. ун-ту внутрішніх справ. Серія юридична. — 2015. — № 2. — С. 285–291.
4. *Кумченко Ю.О.* Інформаційна технологія ідентифікації персоналу на основі комплексу біометричних параметрів : дис. … канд. техн. наук: 05.13.06 / Ю.О. Кумченко. — Херсон, 2017. — 129 с.
5. *Мясіщев О.* Голосове керування віддаленими пристроями через мережу інтернет / О. Мясіщев, І. Муляр // Зб. наук. пр. Військ. ін-ту Київ. нац. ун-ту імені Тараса Шевченка. — 2017. — № 55. — С. 62–71.
6. *Щербаков Є.Ю.* Застосування математичних моделей для голосової ідентифікації суб'єктів у сфері фінансової безпеки / Є.Ю. Щербаков // Нейронечіткі технології моделювання в економіці. — 2017. — № 6. — С. 158–190.
7. *Shah H.N.M.* Biometric Voice Recognition in Security System / H.N.M. Shah, M.Z. Ab Rashid // Indian Journal of Science and Technology. — 2014. — Vol. 7, N 1. — P. 104–112.
8. *An Overview* and Analysis of Voice Authentication Methods. — Available at: https://www.semanticscholar.org/paper/An-Overview-and-Analysis-of-Voice-

Authentication-Shoup-Talkar/572af444f0382b8e7e156ab36192da95a3b8dec4 (дата звернення: 11.11.2019).

9. *Dejavu:* Audio Fingerprinting and Recognition in Python. Available at: https://github.com/ worldveil/dejavu (дата звернення: 11.11.2019).
10. *Martinez J.* Speaker recognition using Mel frequency Cepstral Coefficients (MFCC) and Vector quantization (VQ) techniques / J. Martinez, H. Perez, E. Escamilla // CONIELECOMP 2012, 22nd International Conference on Electrical Communications and Computers. — 2012. — N 1. — P. 248–251. — DOI: 10.1109/CONIELECOMP.2012.6189918
11. *Kelly A.* The Effects of Windowing on the Calculation of MFCCs for Different Types of Speech Sounds / A. Kelly, C. Gobl // Advances in Nonlinear Speech Processing. NOLISP 2011. — Vol. 7015. — 2011.
12. *Welcome* to python_speech_features's documentation! — Available at: https://python-speech-features.readthedocs.io/en/latest/ (дата звернення: 11.11. 2019).
13. *Mel* frequency cepstral coefficient (mfcc) tutorial.— Available at: http://www.practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/ (дата звернення: 11.11.2019).
14. *Open* Speech and Language Resources. — Available at: http://www.openslr.org/12 (дата звернення: 11.11.2019).

Надійшла 12.11.2019