УДК 004.42

ВИКОРИСТАННЯ АЛГОРИТМІВ NAIVE BAYES, RANDOM FOREST TA SVM ДЛЯ КЛАСИФІКАЦІЇ ЗАХВОРЮВАНЬ

Котик О.В., науковий керівник Семко В.В.

Задача автоматичного перетворення мовних повідомлень у текстову форму є однією з ключових у сфері комп'ютерної лінгвістики та машинного навчання. Вона застосовується в голосових інтерфейсах, системах диктування, медичних транскрипціях, а також у розробці доступних технологій для людей з порушенням слуху. У даній роботі досліджуються два підходи до розпізнавання мовлення — нейронна модель Whisper та класичний статистичний підхід на основі чистої прихованої марковської моделі (HMM).

Приховані марковські моделі (HMM) — це один із найперших і найпоширеніших підходів до моделювання мовних послідовностей [1]. У своїй класичній реалізації чиста HMM не поєднується з іншими моделями (наприклад, GMM або DNN), а виконує розпізнавання лише на основі статистичного аналізу послідовностей спостережень.

У моделі визначаються стани (відповідні фонемам або частинам мови), імовірності переходу між ними та імовірності появи ознак для кожного стану. Такий підхід дозволяє точно моделювати часову структуру мовлення, однак має обмежену точність, особливо при наявності шуму або нестандартної дикції.

У процесі дослідження було реалізовано HMM-модель, яка розпізнавала українські мовні повідомлення на основі MFCC-ознак. Навчання здійснювалось на заздалегідь зібраному корпусі аудіофраз, після чого перевірялась точність розпізнавання нових повідомлень. Незважаючи на обмеженість моделі, вона показала задовільні результати при роботі з короткими командами (точність ~78%) [2].

Whisper — це нейронна модель, яка виконує багатомовне розпізнавання мовлення, переклад та транскрипцію. Архітектура моделі базується на трансформерах, які дають змогу обробляти аудіо-сигнал у вигляді спектрограм і виявляти закономірності в контексті фраз [3]. Whisper навчене на великій кількості реальних записів, включаючи дані з шумами, що забезпечує її високу стійкість у реальних умовах.

У рамках роботи було протестовано open-source реалізацію Whisper з використанням українських та англійських аудіофраз. Модель демонструє точність трансформації понад 90% навіть при середньому рівні фонових шумів. Система не потребує ручного створення мовної моделі, побудови словника чи фонематичних транскрипцій [4].

Порівняльний аналіз показав такі результати:

1. HMM добре працює в умовах контрольованих даних, з невеликим набором фіксованих фраз, але втрачає точність при зміні темпу мовлення або при появі шумів (до 20% помилок) [2].

2. Whisper забезпечує стабільне розпізнавання в широкому діапазоні умов, легко адаптується до різних мов, стилів мовлення та акцентів [3, 4]. Також Whisper не вимагає ручного налаштування транскрипцій чи створення фонематичних словників, що значно спрощує реалізацію.

Чиста HMM-модель має обмеження за точністю, однак дозволяє повністю контролювати процес розпізнавання та має високу пояснюваність. Її доцільно використовувати в задачах зі сталою структурою вхідних повідомлень.

Whisper — потужний інструмент, що забезпечує високу якість розпізнавання мовлення «із коробки», навіть у складних умовах.

З урахуванням результатів, Whisper рекомендовано для задач широкого застосування, тоді як HMM може бути корисною для освітніх цілей, демонстрацій роботи статистичних моделей, або як частина легких систем із низькими вимогами до обчислень.

Крім точності, важливим критерієм була швидкість обробки та обчислювальна складність. HMM системи можуть бути менш вимогливими до ресурсів, проте потребують складного налаштування та ручної обробки корпусів. Whisper натомість має просту інтеграцію та дозволяє отримати якісну транскрипцію без значного попереднього налаштування.

У прикладних задачах (онлайн-транскрипція, голосові асистенти) модель Whisper виявилась більш придатною через високу точність, підтримку багатьох мов та стійкість до шуму.

Подальші дослідження можуть бути зосереджені на гібридних системах, що поєднують сильні сторони HMM та нейронних моделей, а також на оптимізації нейромережевих рішень для роботи на пристроях із обмеженими ресурсами.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Rabiner, L. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE.
2. Jurafsky, D., & Martin, J. H. (2022). Speech and Language Processing. 3rd edition draft.
3. OpenAI. (2022). Whisper: Robust Speech Recognition via Large-Scale Weak Supervision. [Електронний ресуср] – Режим доступу - https://openai.com/research/whisper
4. Radford, A., et al. (2022). Whisper: Multilingual speech recognition with a transformer model. arXiv preprint arXiv:2212.04356.

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ БІОРЕСУРСІВ І ПРИРОДОКОРИСТУВАННЯ УКРАЇНИ
ФАКУЛЬТЕТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ

ТЕОРЕТИЧНІ ТА ПРИКЛАДНІ АСПЕКТИ РОЗРОБКИ КОМП'ЮТЕРНИХ СИСТЕМ

ЗБІРНИК НАУКОВИХ ПРАЦЬ ЗА МАТЕРІАЛАМИ
VII ВСЕУКРАЇНСЬКОЇ НАУКОВО-ПРАКТИЧНОЇ КОНФЕРЕНЦІЇ
СТУДЕНТІВ І АСПІРАНТІВ
24 квітня 2025 року

Київ 2025

УДК 004

Відповідальна за випуск: М.І. Лендєл

Збірник наукових праць за матеріалами VII Всеукраїнської науково-практичної конференції студентів і аспірантів «ТЕОРЕТИЧНІ ТА ПРИКЛАДНІ АСПЕКТИ РОЗРОБКИ КОМП'ЮТЕРНИХ СИСТЕМ '2025», 24 квітня 2025 року, НУБіП України, Київ. – 452 с. (електронне видання)

Відповідальність за зміст публікацій несуть автори.