Кундій Дмитро Віталійович, магістр

Чернівецький національний університет імені Ю. Федьковича, Чернівці

**СИСТЕМА РОЗПІЗНАВАННЯ ТА ОБРОБКИ ГОЛОСОВИХ КОМАНД З ВИКОРИСТАННЯМ ЗАСОБІВ KALDI**

Однією з найважливіших задач в області обробки мовлення є розпізнавання та обробка голосових команд, які можуть бути використані в різних додатках, як-от віртуальні асистенти, системи навігації та інші. Для вирішення цієї задачі можна використовувати різні інструменти, серед яких особливе місце займає Kaldi [1].

Kaldi – це відкрите програмне забезпечення, призначене для розпізнавання та обробки мовлення. Цей засіб розроблений у Колорадському університеті та використовується для розпізнавання мовлення різними мовами, зокрема українською [2].

Kaldi використовує комплексний підхід до розпізнавання мовлення, який охоплює такі етапи, як попередній аналіз сигналу, вилучення ознак, моделювання, декодування та пост-обробку результатів [3]. Для кожного з цих етапів Kaldi використовує відповідні алгоритми та методи, до яких належать:

1. Попередній аналіз сигналу: на цьому етапі сигнал аналізується з метою вилучення корисної інформації. Для цього використовуються такі методи, як FFT (швидке перетворення Фур'є), фільтрація та нормалізація сигналу.
2. Вилучення ознак: цей етап передбачає вилучення з сигналу ознак, які можуть бути використані для подальшого розпізнавання мовлення. Для цього використовуються такі методи, як MFCC (коефіцієнти мел-частотної шкали), PLP (параметризація лінійного передавального шляху) тощо.
3. Моделювання: на цьому етапі створюються моделі, які можуть розпізнавати мовлення. Для цього використовуються такі методи, як HMM (приховані моделі Маркова), DNN (глибокі нейронні мережі), GMM (змішані гаусівські моделі) та інші.
4. Декодування: відбувається розпізнавання мовлення за допомогою моделей, створених на попередньому етапі. Для цього використовуються такі алгоритми, як Viterbi decoding та алгоритми пошуку найкращого шляху.
5. Пост-обробка результатів: на цьому етапі виконується підтвердження та корекція результатів розпізнавання, а також виконуються інші завдання, до яких належать розпізнавання іменованих сутностей та визначення мовленнєвого контексту.

На рис. 1 зображено модель розпізнавання мовлення на основі інструментарію Kaldi [4].

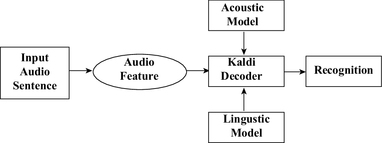


Рис.1. Модель розпізнавання мовлення

Порівняно з іншими системами розпізнавання мовлення, Kaldi відомий своєю високою точністю та ефективністю, які досягаються завдяки використанню сучасних алгоритмів та методів, а також застосуванню різних оптимізаційних технік, як-от обчислювальні графи та масова обробка даних [5].

Метою створення додатку для розпізнавання тексту, що базується на Kaldi, є полегшення та прискорення роботи в різних сферах, пов'язаних із розпізнаванням мовлення. До того ж такий додаток може допомогти зменшити витрати на ручне введення тексту та організувати взаємодію з комп’ютером безпосередньо голосом. Також створення такого додатку має важливе значення для розвитку технологій штучного інтелекту, оскільки він може бути використаний у різноманітних дослідженнях та експериментах, пов'язаних з розпізнаванням мовлення та обробкою текстових даних [6].

Для розробки додатку, який базуються на Kaldi, було обрано платформу Qt. Qt – це кросплатформна бібліотека для розробки програмного забезпечення мовами програмування C++ та QML, яка надає можливості для розробки інтерфейсів користувача, мережевого програмування, роботи з базами даних та ін [7].

Використання Qt для розробки додатків, що базуються на Kaldi, дає змогу забезпечити кросплатформовість та зручний графічний інтерфейс користувача. До того ж використання C++ для розробки додатку є запорукою високої продуктивності та ефективності обробки даних [8].

Результатом роботи над проектом, що базується на Kaldi для розпізнавання тексту, є функціональний кросплатформенний продукт, який забезпечує управління комп’ютером за допомогою голосових команд.

**Література**

1. Y. Lei, L. Xie, R. Singh, and Y. Liu, "Automatic Speech Recognition - A Brief Survey," IEEE Internet of Things Journal: [Електронний ресурс] – Режим доступу: https://ieeexplore.ieee.org/document/8127315.
2. Kaldi офіційна документація: [Електронний ресурс] – Режим доступу: https://kaldi-asr.org/doc/.
3. D. Povey, A. Ghoshal, G. Boulianne, L. Burget, O. Glembek, N. Goel, M. Hannemann, P. Motlicek, Y. Qian, P. Schwarz, J. Silovsky, G. Stemmer, and K. Vesely, "The Kaldi Speech Recognition Toolkit," IEEE 2011 Workshop on Automatic Speech Recognition and Understanding, 2011.
4. Continuous Punjabi speech recognition model based on Kaldi ASR toolkit: [Електронний ресурс] – Режим доступу: https://link.springer.com/article/10.1007/s10772-018-9497-6
5. D. King, "Automatic Speech Recognition Using Deep Neural Networks and Finite State Transducers," Ph.D. thesis, University of Sheffield, 2014.
6. L. R. Rabiner and B. H. Juang, "Fundamentals of Speech Recognition," Prentice-Hall, 1993.
7. Офіційна документація Qt: [Електронний ресурс] – Режим доступу: <https://doc.qt.io/qt-5/>
8. Сайт Cplusplus.com: [Електронний ресурс] – Режим доступу: <https://www.cplusplus.com/> .