6. «Sage Days 29 talk: Robert Bradshaw – Cython.» [Електронний ресурс]. – Режим доступу: https://www.youtube.com/watch?v=osjSS2Rrvm0

7. Документація Cython. Faster code via static typing. [Електронний ресурс]. – Режим доступу: http://docs.cython.org/src/quickstart/cythonize.html

8. Часті питання по Cython. Is Cython faster than CPython? [Електронний ресурс]. – Режим доступу: https://github.com/cython/cython/wiki/FAQ#is-cython-faster-than-cpython

9. PyPy Project. What is PyPy [Електронний ресурс]. – Режим доступу: http://pypy.org/

10. Документація RPython. [Електронний ресурс]. – Режим доступу: https://rpython.readthedocs.org/en/latest/

11. C. F. Bolz, A. Cuni, M. Fijalkowski, M. Leuschel, S. Pedroni, andA. Rigo, «Runtime Feedback in a Meta-tracing JIT for Efficient DynamicLanguages» [Електронний ресурс]. – Режим доступу: http://doi.acm.org/10.1145/2069172.2069181

Андрусенко В.В.

студент,

Харківський національний університет радіоелектроніки

РОЗПІЗНАВАННЯ МОВИ. ВИДІЛЕННЯ ІНФОРМАЦІЙНИХ ОЗНАК МЕТОДОМ МЕЛ-ЧАСТОТНИХ КЕПСТРАЛЬНИХ КОЕФІЦІЄНТІВ

Розпізнавання мови є однією з найпоширеніших задач у галузі машинного навчання на сьогоднішній день. Це, перш за все, зумовлено стрімко зростаючими потребами людства до інтерфейсів взаємодії між людиною та електротехнікою, а також необхідністю впровадження автоматизованих рішень у різноманітних сферах діяльності з метою мінімізації часових та матеріальних витрат. Прикладами використання автоматизованих систем розпізнавання мови у реальному часі можуть бути системи розпізнавання у колл-центрах, які допомагають надавати консультації щодо базових питань та обслуговувати більшу кількість користувачів без необхідності у збільшенні чисельності персоналу, або ж такі сервіси, як Google Voice та Siri, що надають можливість голосового пошуку. Різноманітні рішення до задачі розпізнавання мови розробляються вже протягом багатьох років. У результаті цього була розроблена загальна структура системи автоматичного розпізнавання мови, яка зображена на рисунку 1. Зокрема у межах даної роботи буде більш детально розглянуто таку її складову, як виділення інформаційних ознак [1] та одну з її реалізацій – мел-частотні кепстральні коефіцієнти [2].

Виділення інформаційних ознак з акустичного сигналу є ключовим етапом розпізнавання, адже саме його вихідні дані використовуються для тренування акустичних моделей, наприклад, нейронної мережі або прихованої моделі Маркова, а також для встановлення відповідності між вхідними даними та існуючими у тренувальній базі зразками. Серед найбільш поширених методів виділення ознак виділяють наступні: кодування з лінійним предиктором, перцепційне лінійне передбачення та мел-частотні кепстральні коефіцієнти.

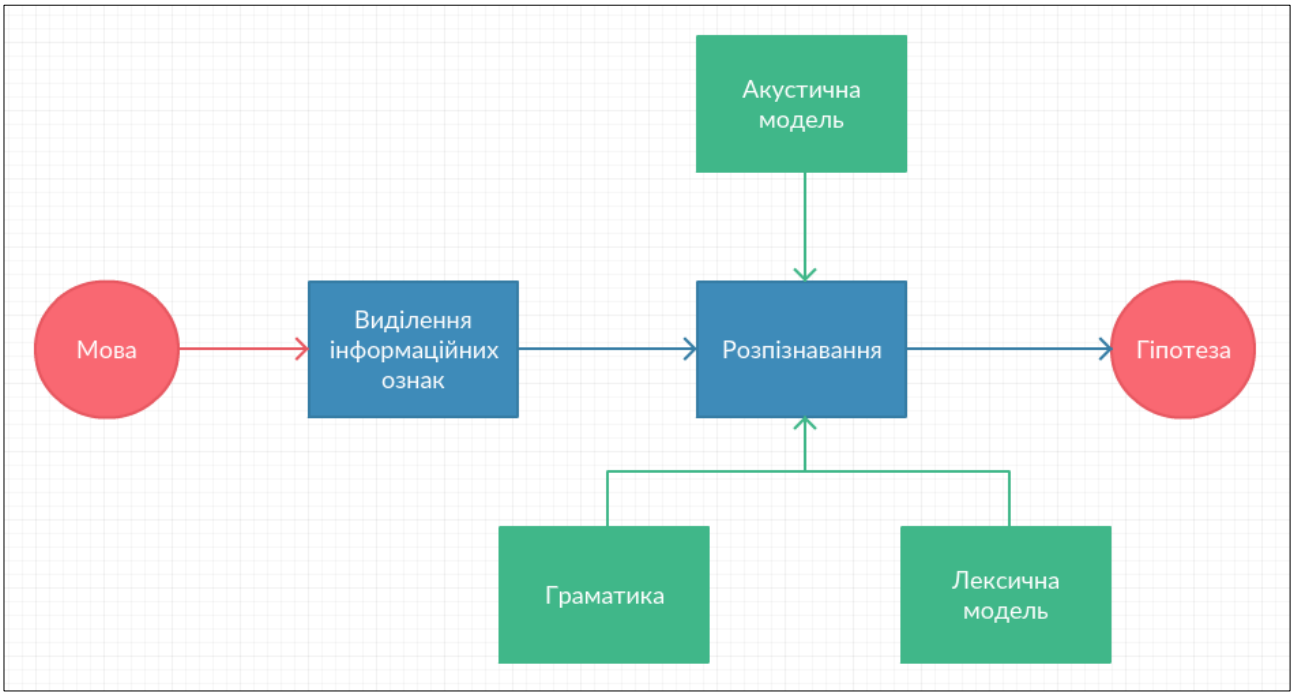

Рис. 1. Узагальнена структура системи автоматичного розпізнавання мови

У залежності від конкретного алгоритму, виділення інформаційних ознак дозволяє одночасно вирішити декілька ключових проблем пов'язаних із розпізнаванням мови, а саме:

а) фоновий шум – проблема, яка можем бути зумовлена як довколишнім середовищем, так і самим спікером;

б) різноманіття спікерів – проблема, що полягає у різній вимові певної послідовності різними людьми у залежності від їх віку, статі, інтонації, діалекту, різного роду анатомічних особливостей та інших індивідуальних ознак;

в) зовнішні фактори: положення мікрофона по відношенню до спікера, напрямок, у якому направлено мікрофон, та інші;

г) об'єм тренувальних даних – кількість пам'яті, що необхідна для зберігання тренувальної бази, на основі якої і виконується розпізнавання.

Одним з найбільш поширених методів виділення інформаційних ознак є мел-частотні кепстральні коефіцієнти (Mel-Frequency Cepstral Coefficients), які по суті є відображенням енергії звукового сигналу. Свою популярність даний метод здобув завдяки тому, що дозволяє одночасно вирішити весь перелік зазначених вище проблем розпізнавання мови у реальному часі. Алгоритм базується на частотному спектрі, який вважається значно точнішим за часовий, із його подальшим відображенням на мел-шкалу. Мел – психофізична одиниця висоти звуку, основою якої є статистична обробка значної кількості даних щодо суб'єктивного сприйняття висоти звуку. Таким чином відображення на цю шкалу надає можливість виділяти найбільш значущі з точки зору сприйняття людиною частоти. Для конвертації герц у мели та навпаки використовуються наступні формули:

$$m = 1127 \ln(1 + \frac{f}{700}),$$
$$f = 700 * \left(e^{\frac{m}{1127}} - 1\right). \quad (1)$$

Одиницею вхідних даних для обробки даним алгоритмом є фрейм – відрізок вхідного сигналу, що відповідає невеликому часовому проміжку, зазвичай – один або декілька десятків мілісекунд. Використання фрейму, як одиниці даних для обробки, дозволяє отримати більш інформативні результати аналізу звукової хвилі у порівнянні із окремим значенням сигналу. Також, для покращення результатів аналізу, сигнал розбивають на фрейми таким чином, що кожен наступний фрейм перетинається із попереднім, а не починається з його кінця. Загальна послідовність отримання мел-частотних кепстральних коефіцієнтів із фрейму зображена на рисунку 2.

Рис. 2. Алгоритм отримання мел-частотних кепстральних коефіцієнтів

Першим етапом обробки окремого фрейму є застосування до нього віконної функції з метою згладжування фрейму на початку і в кінці, що в свою чергу покращить спроможність перетворення Фур'є до вилучення спектральних даних. Зазвичай у даному алгоритмі використовується вікно Хемінга:

$$w(n) = 0.54 - 0.46 \cos(\frac{2\pi n}{N-1}), \quad (2)$$

де N – розмір фрейму; n – індекс поточного значення фрейму.

Наступним до фрейму застосовується перетворення Фур'є. Дане перетворення дозволяє отримати частотні складові звукового сигналу. З метою підвищення швидкості роботи алгоритму можна застосувати швидку реалізацію цього перетворення. Формула дискретного перетворення Фур'є має наступний вигляд:

$$X_k = \sum_{n=0}^{N-1} x_n e^{\frac{-2\pi i}{N} kn}, \quad (3)$$

де N – розмір фрейму;

n – індекс поточного значення фрейму;

x_n – поточне значення фрейму;

k – індекс поточної частоти.

Далі до отриманих на попередньому кроці частот застосовується цифровий фільтр [3] з метою підвищення значущості певних частот. Безпосередньо для даного алгоритму використовується мел-частотний фільтр, який зробить низькі частоти більш вагомими у порівнянні з високими, відображаючи таким чином особливості слуху людини. Фільтри розраховуються наступним чином:

$$f_i = \frac{floor(N+1) * h_i}{samplerate},$$

$$H_m(k) = \begin{cases} 0 & k < f_{m-1} \\ \frac{k - f_{m-1}}{f_m - f_{m-1}} & f_{m-1} \le k \le f_m \\ \frac{f_{m+1} - k}{f_{m+1} - f_m} & f_m \le k \le f_{m+1} \\ 0 & k > f_{m+1} \end{cases}, \quad (4)$$

де N – розмір фрейму;

h_i – опорна точка мел-шкали в герцах;

f_i – опорна точка мел-шкали відносно спектра фрейму;

m – індекс поточного фільтру;

k – індекс поточного значення фрейму.

Для зменшення чутливості кінцевих значень мел-частотних кепстральних коефіцієнтів до шумів фільтри необхідно застосовувати до енергії спектра сигналу, а не його безпосередніх значень, а від отриманого результату взяти логарифм:

$$S_m = \log \sum_{k=0}^{N-1} |X[k]|^2 * H_m[k], \quad (5)$$

де $X[k]$ – поточне значення фрейму;

$H_m[k]$ – поточне значення фільтру з індексом m.

Останнім кроком є стиснення отриманих коефіцієнтів за допомогою дискретного косинусного перетворення з метою надання більшої ваги першим коефіцієнтам, що представляють форму спектра, у порівнянні з останніми, що здебільшого відображають шуми:

$$C_l = \sum_{m=0}^{M-1} S_m * \cos\left(\frac{\pi * l * (m + 0.5)}{M}\right), \quad (6)$$

де m – індекс поточного значення відфільтрованої енергії спектра,

l – індекс кепстрального коефіцієнта, що обчислюється.

Отриманий набір коефіцієнтів називається акустичним вектором, що включає в себе найбільш важливі з точки зору фонетики характеристики мови, які в свою чергу можуть бути використані для подальшого аналізу, наприклад, для тренування акустичних моделей на базі нейронної мережі або прихованої марковської моделі.

Список використаних джерел:

1. Urmila Shrawankar, Dr. Vilas Thakare «Techniques for Feature Extraction in Speech Recognition System: A Comparative Study», International Journal of Computer Applications in Engineering, Technology and Sciences, pp. 412-418, 2013.

2. Namrata Dave «Feature Extraction Methods LPC, PLP and MFCC in Speech Recognition», International journal for advance research in engineering and technology, vol. 1, no. 6, 2013.

3. Ben J. Shannon, Kuldip K. Paliwal «A Comparative Study of Filter Bank Spacing for Speech Recognition», Microelectronic engineering research conference, 2003.

4. Anant G. Kulkarni, Dr. M. F. Qureshi, Dr. Manoj Jha «Discrete Fourier Transform: Approach to Signal Processing», International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, vol.3, no. 10, 2014.