SECTION 12.
COMPUTER AND SOFTWARE ENGINEERING

Ялова Катерина Миколаївна
канд, техн. наук, доцент, зав. кафедри програмного забезпечення систем
Дніпровський державний технічний університет, Україна

Шелюг Костянтин Юрійович
аспірант кафедри програмного забезпечення систем
Дніпровський державний технічний університет, Україна

МЕТОДИ РОЗПІЗНАВАННЯ МОВЛЕННЯ

Питання підвищення якості людино-машинної взаємодії є актуальним науково-практичним завданням. Найбільш ефективними способами взаємодії людини з машиною є ті, які є природними для людини: через візуальні образи і мовлення. Мова є найбільш природною формою людського спілкування і тому реалізація інтерфейсу на основі аналізу мовленнєвої інформації є перспективним напрямком розвитку інтелектуальних систем управління. Створення мовленнєвих інтерфейсів може знайти застосування в системах різного призначення: голосове управління для людей з обмеженими можливостями, автовідповідачі, оброблення дзвінків в автоматичному режимі, ідентифікація власника пристрою, голосового керування «розумним домом» тощо. Реалізація мовленнєвого інтерфейсу базується на системі автоматичного розпізнавання мови, що забезпечує прийом мовленнєвого сигналу і перетворення його до тексту або команди. Система автоматичного розпізнавання мовлення (Automatic Speech Recognition System – ASRS) [1] – це інформаційна система, що перетворює вхідний мовленнєвий сигнал в розпізнане повідомлення. При цьому розпізнане повідомлення може бути представлено як у формі тексту, так і одразу перетворено в зручну для його подальшої обробки форму. Більшість систем розпізнавання мовлення складається з процесу аналізу й обробки аналогового сигналу і процесу розпізнавання. При аналізі аналогового сигналу з промови виділяються властивості, які в подальшому використовуються в процесі розпізнавання для того, щоб визначити, що було сказано. Системи автоматичного розпізнавання мовлення класифікуються за такими ознаками:

- розмір словника (обмежений набір слів або великий словник);
- залежність від диктора (дикторозалежні або дикторонезалежні);
- тип мовлення (злітне, роздільне);
- призначення (системи диктування, командні системи);
- алгоритм розпізнавання, що використовується;
- тип структурної одиниці (фрази, слова, фонеми тощо);
- принципи виділення структурних одиниць.

Загальна схема процесу розпізнавання мовленнєвого сигналу наступна: отримання акустичного сигналу, який поступає з мікрофону користувача (або завантаження аудіо файлу), оцифровування звукового сигналу, отримання характеристик сигналу та порівняння вектору ознак вхідного сигналу з шаблонами. Традиційно завдання розпізнавання мовлення поділяють на:

1. Завдання ідентифікації – це завдання верифікації та ідентифікації дикторів.

2. Завдання класифікації – це завдання розпізнавання ключових слів, розпізнавання зв'язного мовлення і завдання семантичного аналізу мови.

3. Завдання діагностики – це завдання визначення психофізичного стану диктора.

Не дивлячись на стрімко зростаючі обчислювальні потужності, створення систем розпізнавання мовлення залишається надзвичайно складною проблемою, що обумовлено її міждисциплінарним характером (необхідно володіти знаннями з філології, лінгвістики, цифрової обробки сигналів, акустики, статистики, розпізнавання образів тощо), так і високою обчислювальною складністю розроблених алгоритмів.

Найперші спроби створення систем автоматичного розпізнавання мовлення почалися в 1950-х роках, коли була розроблена перша залежна від диктора система, що розпізнавала цифри, в якості характеристик вхідного сигналу використовувалися резонанси голосних звуків в словах. У 70-х роках було відкрито метод Dynamic Time Warping – DTW і метод кодування лінійного передбачення (Linear Predictive Coding – LPC), що використовуються до тепер. Найсучаснішими методами розпізнавання мовлення є нейромережеві методи. Методи розпізнавання мовлення можна розділити на такі види:

1. Приховані Марківські моделі – моделі, які складаються з N станів, де деяка систем може приймати одне з M значень певного параметра.

2. Методи лінійного передбачення, які використовують для обробки звукових сигналів для представлення спектру цифрового сигналу мови в стиснутому вигляді, використовуючи дані лінійної моделі прогнозування.

3. Часові динамічні методи, які полягають в розпізнаванні мовлення на основі процедури порівняння розпізнаного фрагменту із шаблоном. Методи дозволяють знайти різницю між двома рядами мовленнєвих команд.

4. Методи засновані на використанні нейронних мереж різноманітної архітектури [2]. Розпізнавання мовлення з дискретним спостереженням найчастіше реалізується із використанням одношарових перцептронів. Для векторного квантування можуть бути застосовані нейронні мережі подібні до карт ознак Кохонена. Багатошарові нейронні мережі можуть бути використані для зниження розмірності векторів ознак, які формуються на початковому етапі розпізнавання.

Висновки. Вибір конкретного методі розпізнавання, що буде реалізовано в системі розпізнавання мовлення, суттєво залежить від характеристики системи автоматичного розпізнавання мовлення та кінцевої мети її застосування.

Список використаних джерел:

1. Alharbi S., Alrazgan M., & Alrashed A. (2021). Automatic Speech Recognition: Systematic Literature Review. IEEE Access, (9), 131858-131876. https://doi: 10.1109/ACCESS.2021.3112535.
2. Papstratis I. (2022). Speech recognition: a review of the different deep learning approaches. Accessed https://theaisummer.com/speech-recognition/ .