УДК 004.852

Марчук Г.В., ст. викладач
Любченко Д.В,. асистент
Державний університет «Житомирська політехніка»

ПІДВИЩЕННЯ ТОЧНОСТІ ASR ДЛЯ OOV-ЛЕКСИКИ

Розпізнавання мови (ASR) є ключовою технологією в сучасній людино-машинній взаємодії. Фреймворк Speech в iOS забезпечує доступ до потужних нейромережевих моделей, які ефективно працюють із загальновживаною лексикою, що тренована на великих мовних масивах [1]. Точність стандартних ASR-систем критично знижується при роботі з вузькоспеціалізованою лексикою, відомою як Out-of-Vocabulary (OOV) слова. Це стосується медичної термінології, назв фармацевтичних препаратів, технічного жаргону та унікальних власних назв. Недостатня точність у критично важливих галузях, як-от охорона здоров'я, може призвести до серйозних помилок, що обумовлює нагальну потребу в механізмах спеціалізації мовної моделі без повного перенавчання базового ASR-рушія.

Процес автоматичного розпізнавання мовлення (ASR) є, по суті, задачею максимізації ймовірності. Мета - знайти та вивести найбільш імовірну послідовність слів (W_{final}) за умови отриманого акустичного спостереження або звукового сигналу (O). Ця фундаментальна задача розв'язується за допомогою Теореми Баєса, яка дозволяє декомпозицію складної умовної ймовірності $P(W|O)$ на два основні, незалежно керовані, компоненти:

$$\hat{W} = \arg\max_{W} P(O \mid W) \cdot P(W), \qquad (1)$$

де $P(O|W)$ (Акустична модель) - оцінює ймовірність того, що саме послідовність слів W була вимовлена, призвівши до отриманого акустичного сигналу O; P(W) (Мовна модель, LM) - оцінює апріорну (попередню) ймовірність виникнення послідовності слів W у природній мові. Цей компонент відповідає за синтаксичну та семантичну коректність, виступаючи як потужний фільтр, що відсікає малоймовірні або безглузді послідовності.

Запропонований підхід інтеграції кастомних словників з явним прописом фонем є ефективним та необхідним механізмом спеціалізації ASR-систем в екосистемі Apple. Цей метод долає обмеження, притаманні базовим моделям, забезпечуючи клінічно або функціонально необхідний рівень точності розпізнавання OOV-слів. Це є критично важливим для створення надійних голосових інтерфейсів у

високоспеціалізованих галузях (охорона здоров'я, юридична сфера, технічні системи), де неточність розпізнавання є неприпустимою.

Базові мовні моделі ($P_{base}(W)$) тренуються на дуже великих масивах загальновживаного тексту. Якщо користувач вимовляє вузькоспеціалізований термін (наприклад, медичний діагноз, назва бренду, технічний термін), який не міститься у тренувальних даних базової моделі (так зване OOV – Out-Of-Vocabulary слово, W_{custom}), то ймовірність $P_{base}(W_{custom})$ буде надзвичайно низькою, часто близькою до нуля. Низька ймовірність переважає навіть сильний сигнал акустичної моделі, що призводить до помилки розпізнавання і заміни спеціалізованого слова на якесь фонетично схоже, але загальновживане слово.

Для вирішення цієї проблеми застосовується модель інтерполяції. Це процес злиття (fusion) двох мовних моделей - загальної та спеціалізованої - за допомогою зваженого середнього.

Інтерполяція коригує апріорну ймовірність P(W) саме для цільових спеціалізованих термінів. Кінцева (інтерпольована) ймовірність Pfinal(W) розраховується за формулою:

$$P_{final}(W) = (1-\lambda)P_{base}(W) + \lambda P_{custom}(W), \quad (2)$$

де $P_{base}(W)$ - ймовірність базової моделі отримана зі стандартної, великої, але неспеціалізованої мовної моделі. Для слів W_{custom} ця величина практично дорівнює 0. $P_{custom}(W)$ - ймовірність кастомної моделі, яка спеціально сконструйована і містить кастомний словник. Для цільових слів W_{custom} ця ймовірність штучно встановлюється як висока (наприклад, рівномірно розподілена для всіх слів словника). λ - коефіцієнт ваги інтерполяції, який є ключовим керуючим параметром, де $\lambda \in [0, 1]$. Якщо $\lambda = 0$, використовується лише базова модель, а якщо $\lambda = 1$, використовується лише кастомний словник. Оптимальне значення λ визначає ступінь довіри системи до спеціалізованого словника.

Завдяки інтерполяції, для спеціалізованих термінів W_{custom}, навіть невеликий, але ненульовий коефіцієнт λ призводить до значного збільшення кінцевої ймовірності $P_{final}(W_{custom})$.

Список використаних джерел:

1. Apple Developer Documentation. Speech Framework [Електронний ресурс]. 2025. URL: https://developer.apple.com/documentation/Speech (дата звернення: 30.10.2025).