

ПРЕДСТАВЛЕНИЯ СЛОВНЫХ СЕТЕЙ В ЗАДАЧЕ ИНДЕКСИРОВАННОГО ПОИСКА ПО РЕЧЕВЫМ ДАННЫМ

Петров О. Е. (Университет ИТМО), Кабаров В. И. (Университет ИТМО)
Научный руководитель – д.т.н. Матвеев Ю. Н.
(Университет ИТМО)

В работе описываются различные способы представления словных сетей, полученных в процессе автоматического распознавания речи, для построения обратных индексов поиска [1]. Такие структуры позволяют с одной стороны сохранить множество порожденных в процессе лингвистического декодирования гипотез, а с другой — осуществлять быстрый доступ к произвольному узлу сети для построения индекса.

В контакт-центрах и офисах продаж крупных компаний работают тысячи сотрудников, которые ежедневно обрабатывают сотни тысяч обращений. В целях контроля качества обслуживания все разговоры записываются, формируя большие объемы аудиоданных, которые хранятся в постоянно пополняющихся архивах.

Современные системы автоматического распознавания речи позволяют получать текстовое представление диалога клиента и оператора. Одним из способов организации поискового индекса по текстам диалогов является использование обратного индекса слов. Применительно к задачам речевой аналитики такой подход позволяет создавать решения, обеспечивающие одновременно и высокую скорость индексации, и быстрый гибкий поиск. Точность поиска при этом ограничивается точностью используемой системы автоматического распознавания речи. Использование текстового представления диалогов клиента и оператора остается невозможным в тех случаях, когда точность распознавания речи недостаточно высокая: малоресурсные языки, диалекты и смеси языков, распознавание в сложных акустических условиях.

Использование словных сетей, получаемых в результате работы лингвистического декодера, позволяет повысить полноту результата за счет возможности обнаружения слов, распознанных с низким уровнем достоверности и не попавших в лучший результат распознавания. Словная сеть представляет собой взвешенный направленный словный граф без циклов. Количество параллельных состояний и переходов из каждого состояния может быть произвольным и зависеть от параметров декодирования и ограничений луча поиска. Такая структура позволяет быстро обходить сеть, но плохо поддается индексации.

Вместо словных сетей в рамках исследования рассматриваются более компактное их представление — сети спутывания и расширенные сети спутывания [2]. Все пути в сети спутывания проходят через все состояния. Набор переходов между парой состояний называется бином. Каждый бин имеет временные метки начала и конца, которые размещаются в соответствующих состояниях. Каждый бин может содержать произвольное количество переходов и не более чем ε -переход, означающий, что на данном бине слово может быть пропущено. Такие переходы используются в том числе для обозначения паузы. Расширенные сети спутывания допускают переходы через несколько состояний, что позволяет значительно сократить количество ε -переходов, но при этом сохранить возможность индексировать гипотезы через позицию бина.

В данной работе рассматривается алгоритм декодирования с минимизацией байесовского риска для построения сетей спутывания с минимальным значением пословной ошибки (Word Error Rate, WER), основанный на статье [3]. В его основе лежит двумерная рекурсия, которая в одном измерении использует алгоритм прямого-обратного хода по словной сети, а в другом — рекурсивный алгоритм вычисления редакционного расстояния Левенштейна.

В рамках работы был создан прототип, реализующий предложенный алгоритм, и получены показатели точности работы системы, подтверждающие возможность

использования его в задачах речевой аналитики для записей диалогов клиента и оператора контакт-центров и офисов продаж.

Список использованных источников:

1. Petrov O. The architecture of a system for full-text search by speech data based on a global search index // *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*. – 2021. – № 5(21). – С. 791–794.
2. Mangu, L., Brill, E., Stolcke, A. Finding consensus in speech recognition: word error minimization and other applications of confusion networks // *Computer Speech & Language* 14.4 – 2000. – С. 373–400.
3. Xu, H., Povey, D., Mangu, L., Zhu, J. Minimum bayes risk decoding and system combination based on a recursion for edit distance // *Computer Speech & Language* 25.4 – 2011. – С. 802–828.