

Тема работы: Быстрая адаптация компонентных систем распознавания речи

Состав коллектива:

1. Научный руководитель 1: Бондаренко Иван Юрьевич
 - место работы: кафедра фундаментальной и прикладной лингвистики НГУ
 - должность: старший преподаватель
 - учёная степень и звание: -
 - адрес электронной почты: i.bondarenko@g.nsu.ru
2. Научный руководитель 2: Батура Татьяна Викторовна
 - место работы: ИСИ СО РАН
 - должность: исполняющий обязанности заведующего лабораторией, старший научный сотрудник
 - учёная степень и звание: кандидат физико-математических наук, доцент
 - адрес электронной почты: t.batura@g.nsu.ru
3. Студент: Гребенкин Даниил Витальевич
 - место учебы: ММФ НГУ, второй курс магистратуры по направлению “Математика и компьютерные науки”
 - адрес электронной почты: d.grebenkin@g.nsu.ru

Научное содержание работы

Постановка задачи и современное состояние проблемы (на момент начала работы). В настоящее время все существующие системы распознавания речи можно различать по структуре: они бывают компонентные и «сквозные». Несмотря на наличия у первых недостатков вроде необходимости настройки каждого модуля, которая требует знаний о системе от пользователя, компонентные системы не требуют переобучения акустико-фонетических модулей, поэтому добавление новых слов в такие системы можно назвать «быстрым». Такая функция крайне востребована специалистами в области обработки естественного языка (NLP) и специалистами в области искусственного интеллекта как при решении задачи адаптации систем распознавания речи широкого профиля, так и при создании систем распознавания специальной лексики, например малоресурсных языков.

Целью данной работы было оценить преимущества и недостатки различных методов быстрой адаптации компонентных систем распознавания речи на уровне языковых моделей на этапах создания гипотез распознавания речи и на этапе их рескоринга другой языковой моделью. Для достижения поставленной цели были решены следующие *задачи*:

1. исследовать и сравнить существующие типы систем распознавания речи, выявить их преимущества и недостатки применительно к задаче адаптации;
2. исследовать методы адаптации компонентных систем распознавания речи на уровне языковых моделей;

3. экспериментально сравнить качество работы компонентной системы распознавания речи с ее модификациями с разными языковыми моделями, адаптированными под определенную предметную область;
4. оценить полученные результаты разных методов адаптации языковых моделей компонентной системы распознавания речи с точки зрения улучшения качества распознавания и количества затраченных вычислительных ресурсов.

Объектом исследования в работе является система распознавания речи на основе компонентного подхода. *Предметом исследования* в работе являются методы компонентных систем распознавания речи.

Подробное описание работы и результаты. В работе были проанализированы существующие типы систем распознавания речи, было определено, что для задачи адаптации наиболее универсальными системами являются компонентные, т.к. возможна отдельная адаптация разных модулей в зависимости от имеющихся видов данных — аудиозаписей или текстов, содержащих лексику из новой предметной области, при этом для адаптации требуются меньшие объемы данных, чем для моделей с end-to-end структурой вследствие наличия регуляризирующих статистических ограничений. Наиболее важным уровнем для адаптации был выбран уровень языковых моделей, т.к.: 1) формирование достаточно больших текстовых корпусов является более ресурсоемкой задачей; 2) сферы применения распознавания речи сильнее отличаются в лексическом смысле, чем в акустико-фонетическом.

Для того, чтобы экспериментально сравнить качество работы компонентной системы распознавания речи с разными языковыми моделями, адаптированными под определенную предметную область, использовались модель распознавания речи `vosk-model-ru-0.22` [] и несколько комбинаций языковых моделей в ее составе:

1. исходные N-граммная модель в качестве основной языковой модели и RNNLM для рескоринга;
2. N-граммная модель в качестве основной языковой модели из модели распознавания речи `vosk-model-ru-0.22` и RuBERT[] для рескоринга;
3. `vosk-model-ru-0.22` и RuBERT для рескоринга.

Применение современных методов адаптации компонентных систем распознавания речи на уровне языковых моделей зависит от типа языковых моделей: интерполяция различных N-граммных или дообучение нейросетевых моделей (BERT, GPT-2, RNN). *Экспериментально было показано*, что оба метода применимы при использовании языковых моделей для формирования наиболее вероятных вариантов распознанного фрагмента речи и решения задачи нахождения лучшего варианта из них (N-best rescoring): различные модификации интерполированной модели с разными типами сглаживания улучшают ее способность к предсказанию следующего слова в цепочке слов тестовой выборки (уменьшение значения перплексии) и повышают качество распознавания слов из новой предметной области без потери качества (уменьшение значений Word Error Rate и Character Error Rate); решение задачи N-best rescoring с помощью «дообучения» нейросетевых моделей типа BERT также улучшает качество распознавания речи, однако использование BERT без

встроенной в vosk-model-ru-0.22 RNNLM показало более низкий результат, хотя слова из новой предметной области были распознаны.

С точки зрения количества затраченных ресурсов для процесса адаптации наименее ресурсоемкими являются нейросетевые модели, т.к. загрузка и интерполяция больших N-граммных моделей требует наличия большого объема оперативной памяти. При ограниченных вычислительных ресурсах компьютера для обучения нейросетевой модели и использования ее для рескоринга можно эффективно использовать параллельные вычисления на GPU при наличии технологии CUDA.

Эффект от использования кластера в достижении целей работы. Ресурсы кластера, в частности Tesla V100, использовались для адаптации N-граммной модели с различными параметрами сглаживания N-грамм, для дообучения RuBERT и для тестирования получившихся моделей.

Перечень публикаций, содержащих результаты работы.

1. Гребенкин Д.В. Быстрая адаптация компонентных систем распознавания речи // Материалы студенческой сессии конференции Диалог-2022, 2022. URL: https://www.dialog-21.ru/media/5803/grebenkin_146.pdf (дата обращения 04.05.2023).