

Министерство образования Республики Беларусь

Учреждение образования
Белорусский государственный университет
информатики и радиоэлектроники

УДК 004.383

Бутома Виталий Сергеевич

Автоматическое построение системы голосового управления

АВТОРЕФЕРАТ

на соискание академической степени
магистра информатики и вычислительной техники

по специальности 1-40 81 04 – Обработка больших объемов информации

Научный руководитель

Волорова Н.А.

кандидат технических наук, доцент

Минск 2019

КРАТКОЕ ВВЕДЕНИЕ

Голосовые ассистенты с каждым годом становятся все популярнее и популярнее. Они принялись быстрее, чем любая другая технология в истории. Столь ощутимый рост популярности голосового поиска можно объяснить заметным улучшением функционала персональных ассистентов и быстрым развитием технологий.

Голосовой поиск развивает почти каждая крупная компания, которая хоть как-то связана с облачными технологиями или мобильными гаджетами. Лидерами рынка долгое время являлись такие продукты, как Siri, Cortana, Google Home, Google Voice, Amazon Echo (Alexa).

По оптимистичным прогнозам, через 10 лет голосовые помощники станут новым способом управления планшетами и компьютерами.

Во-первых, они научатся корректно отвечать на поставленные им вопросы. Уже сегодняшние голосовые помощники не только дают разные ссылки, где можно найти ответ на поставленный вопрос, но и сам ответ.

Во-вторых, разработчики стараются сделать личного помощника совершеннее, превращая его из «пассивного» в «активного». Помощник будет выполнять свои функции прежде, чем вы его попросите. Такое поведение помощника основывается на распознавании вашего поведения, предсказания вашего следующего шага. Помощник быстро станет делом привычки.

Между голосовыми помощниками постоянно будет конкуренция, одни будут справляться с чем-то лучше, с чем-то хуже, однако все чаще можно услышать такую информацию, что с Google и Amazon никто конкурировать не планирует; то есть производители голосовых интерфейсов признают свою неконкурентоспособность в этом плане и шаг за шагом уходят в сторону. Поэтому в данной работе будут рассмотрены проблемы создания логики для голосового управления для таких продуктов, как Amazon Echo (Alexa), Google Home. Настоящая работа призвана показать, что можно разработать механизм голосового управления, который способен конкурировать или дополнить технологии Google и Amazon. Важное замечание, что вся работа в данной статье приведена для английского языка, его лексики и синтаксиса, однако все алгоритмы легко обобщаются и на все остальные языки.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Цель и задачи исследования

Целью диссертационной работы является изучение методов применения нейронных сетей, методов глубокого обучения, для решения задачи построения системы голосового управления по набору начальных ключевых фраз. Разработка алгоритма для определения семантической схожести двух предложений (текстов), алгоритма для генерации парафраз, то есть семантически близких выражений или предложений по исходному. Для достижения поставленной цели необходимо решить следующие задачи:

1. Ознакомиться с методами представления и обработки слов и текстовых данных.
2. Определить сферы применения методов машинного обучения для задач связанными с обработкой текстов, а именно методов,использующие нейронные сети, в том числе глубокого обучения.
3. Разработать алгоритмы для генерации текста, семантически похожего на данный; определения похожести двух предложений (слов), под похожестью имеется в виду семантическая схожесть.
4. Реализовать алгоритмы, реализовать нейронную сеть, собрать необходимые данные, обучить модель.
5. Провести экспериментальные исследования разработанного алгоритма и модели нейронной сети.

Объектом исследования являются голосовые помощники.

Предметом исследования является применение нейронных сетей для обработки предложений, отвечающих за команды.

Основной *гипотезой*, положенной в основу диссертационной работы, является возможность использования глубоких нейронных сетей для таких задач, как машинный перевод, диалоговые модели, порождение текста. Данные задачи относятся к максимальному (третьему) класса задач по сложности среди выделяемых в обработке естественного языка. Для достижения поставленной задачи необходимы данные, которые находятся в открытом доступе в сети интернет. Также пришлось прибегнуть с собственной генерации датасета, так как наборов данных для исследования именно команд не существует.

Личный вклад соискателя

Результаты, приведенные в диссертации, получены соискателем лично. Вклад научного руководителя Н.А.Волорова, заключается в формулировке целей и задач исследования.

Апробация результатов диссертации

Основные положения диссертационной работы докладывались и обсуждались на международной научно-технической конференции молодых ученых «Новые материалы, оборудование и технологии в промышленности» (Могилев, Беларусь, 2018).

Опубликованность результатов диссертации

По теме диссертации опубликовано 2 печатные работ, из них 2 работы в сборниках трудов и материалов конференций.

Структура и объем диссертации

Диссертация состоит из введения, общей характеристики работы, шести глав, заключения, списка использованных источников и приложений. В первой главе представлен анализ предметной области, рассмотрены методы создания логики голосовых интерфейсов а также описаны актуальные проблемы, которые возникают у разработчиков и пользователей при создании голосовой логики. Вторая глава посвящена технологиям, которые использовались во время исследования. В третьей главе происходит небольшой экскурс в рекуррентные нейронные сети, а также в одну из ее модификаций, долгую краткосрочную память. Данные нейронные сети используются в работе, поэтому важно понимание их архитектуры. В четвертой главе описано исследование поставленной задачи, обсуждены основные положения, а также предложены модели НС, которые могут решать поставленные задачи. В шестой главе предложен механизм сбора и генерации данных, а

Общий объем работы составляет 65 страниц, из которых основного текста 52 страницы, 18 рисунков, использованных источников из 22 наименований на 2 страницах и 3 приложения на 10 страницах.

ОСНОВНОЕ СОДЕРЖАНИЕ

Во **введении** определена область и указаны основные направления исследования, показана актуальность темы диссертационной работы, дана краткая характеристика исследуемых вопросов, обозначена практическая ценность работы, кратко затронута тема аналогов.

В **первой главе** проведен анализ таких продуктов как Google Home и Amazon Echo, обозначены минусы и проблемы в создании голосовой логики для данного вида устройств. Также рассмотрены возможные варианты решения существующих проблем.

Во **второй главе** происходит небольшой экскурс в нейронные сети, а также определенного вида архитектуры и адаптации, которые необходимо знать и понимать для успешной реализации данной работы.

В **третьей главе** рассматриваются основные технологии и библиотеки, которые используются в данной работе. В качестве фреймворков для машинного обучения и глубокого обучения использовались Tensorflow и Keras, которые являются наиболее популярными инструментами среди разработчиков и исследователей. Обучение нейронной сети занимает ощутимое время, поэтому была использована технология CUDA для распараллеливания обучения сети на GPU. Далее рассказывается про возможности инструментария NLTK для задач обработки естественного языка, а также про сеть WordNet.

Четвертая глава рассказывает про ход исследования, настройки экспериментов. Были приведены общие положения исследований, какие были выдвинуты гипотезы, которые привели к подобному решению, а также сравнения аналогов.

Были представлены алгоритмы для задач генерации парафразов, генерации предложений по контексту и нахождения семантического расстояния между предложениями.

В **пятой главе** можно найти информацию, алгоритмы для генерации и расширения датасетов для задач, рассматриваемых в данной работе. Особое внимание уделяется генерации своего собственного датасета и объяснению зачем он нужен.

Шестая глава посвящена выводам по данной работе и сравнением результатов с аналогами. Также делаются выводы по дальнейшей работе.

ЗАКЛЮЧЕНИЕ

Основные научные результаты диссертации

В рамках данной работы были изучены и показаны актуальные проблемы создания логики голосовых интерфейсов, указанные проблемы являются камнем преткновения на пути интеграции голосовых интерфейсов в повсеместную жизнь человека.

Для таких проблем, как маппинг наилучшего результата и генерации похожих семантических фраз, а также генерации предложений из слов были предложены успешные решения, которые можно использовать при построении своей системы голосового управления.

Полученные модели готовы составить конкуренцию с самыми современными на данный момент аналогами-моделями.

Открытыми остаются проблемы анаферы, а также вопрос пересечения семантики между сгенерированными фразами. Последние проблемы предлагаются вынести в отдельное исследование, так как они не были целью работы.

Рекомендации по практическому использованию результатов

Полученные результаты формируют теоретическую и практическую базу для разработки ПО для решения задач по улучшению качества работы голосовых интерфейсов. Данные программы могут работать в режиме реального времени и легко интегрироваться как промежуточное звено в системы голосового управления. Также возможно написание отдельных модулей, связывающих Alexa, Google Voice и нашу систему.

Были подтверждены гипотезы про использование объединенных рекуррентных нейронных сетей с остаточными связями для генерации текста, данные результаты могут быть использованы в более крупных исследованиях.

СПИСОК ОПУБЛИКОВАННЫХ РАБОТ

1-А. Бутома В.С., Бутома А.М.: Автоматическое построение системы голосового управления веб-сайтом/ Сборник материалов международной научно-технической конференции молодых ученых "Материалы, оборудование и ресурсосберегающие технологии". – Могилев, 2018. – 389с.

2-А. Духовник А.Н., Бутома В.С., Волорова Н.А.: Метод улучшения результатов сверточных нейронных сетей методом зеркального нахлеста изображения / Сборник материалов международной научно-технической конференции молодых ученых "Новые материалы, оборудование и технологии в промышленности". – Могилев, 2018. – 152 с.