Применение нейросетей для задачи устранения ошибок распознавания речи при редактировании исходного кода программ для ЭВМ с использованием Google Cloud Speech

Д.В. Колчева, А.С. Дмитриев

Волгоградский государственный технический университет

Аннотация: Рассматривается усовершенствование подхода к распознаванию речи для редактирования исходного кода программ для ЭВМ с использованием технологии Google Cloud Speech. В предлагаемом подходе рассмотрено комбинирование использования нейросетей с технологиями обработки звука, редакционными расстояниями и таблицами замены. Представлена архитектура системы для адаптации распознавания выражений на языке Python, рассмотрены достоинства и недостатки такой системы. Приведены результаты анализа данного подхода на прототипе, разработанном для интеграции системы с популярной платформой GitHub, а также мессенджером Telegram.
Ключевые слова: распознавание речи, нейросети, машинное обучение, анализ исходного кода, формальные языки, редакционные расстояния.

Введение

В настоящее время, при разработке приложений стали всё чаще применяться технологии удалённой работы, когда сотрудник находится в месте, удобном ему для работы [1] и при этом может использовать мобильные устройства для решения задач, связанных с программным обеспечением. В случае использования мобильных устройств, ввод и внесение изменений в исходный код программы могут быть затруднены в силу особенностей мобильных устройств из-за недостаточной эргономичности экрана и клавиатуры.

Для решения данной проблемы было разработано веб-приложение Reviewgram [2]. Вышеупомянутое приложение предназначено для внесения правок в исходный код программ, размещённых на платформе GitHub, используя мобильные устройства с интеграцией в платформу Telegram. Приложение поддерживает голосовой ввод правок на языке Python, комбинируя для повышения точности распознавания технологии шумоподавления, редакционные расстояния и таблицы замены, используя тот факт, что распознавание речи позволяет ускорить набор при вводе

коротких команд [3]. Для улучшения эргономичности использовались методы исследования, рассмотренные в статье [4].

Однако при этом возникла задача улучшения исправления ошибок распознавания речи, связанная с особенностью распознавания речи с использованием Google Cloud Speech. Данный инструмент был выбран, так как имеет широкое применение и хорошие результаты по сравнению с другими сервисами распознавания речи [5,6]. Однако, так как данная система имеет целью распознавание естественного языка и при реализации Reviewgram было использовано распознавание английского языка, то при тестировании возникли ситуации, когда в случае неточностей произношения, связанных с русским акцентом, скоростью произношения и другими факторами, система давала недостаточно точный результат, хоть и представляющий собой фрагменты предложений естественного языка, сходные по произношению с требуемыми. Тем не менее результаты состояли из тех же или сходных фонем (звуков речи в языковой системе), что и требуемый результат. Исходя из этой особенности, возникла задача исправления ошибок на основе таких прецедентов и учёта таких ситуаций.

Для решения этой задачи было принято решение использовать нейронные сети в силу их применимости для подобного рода задач [7]. В данном случае прямое использование систем рассуждений, использованных на прецедентах, не рассматривалось из-за потенциальных внутренних зависимостей во входных данных.

Предлагаемый усовершенствованный подход к распознаванию речи для редактирования исходного кода программ для ЭВМ

Для усовершенствования подхода был выбран многослойный перцептрон (далее по тексту указан как классификатор), как довольно простая в реализации и обучении модель нейронной сети [8]. Однако в дальнейшем планируется её замена на более совершенные sequence-to-

sequence модели [9] или модели на основе вариационных автокодировщиков [10]. Начало хода работы нового подхода показано на UML-диаграмме активности на рис. 1.

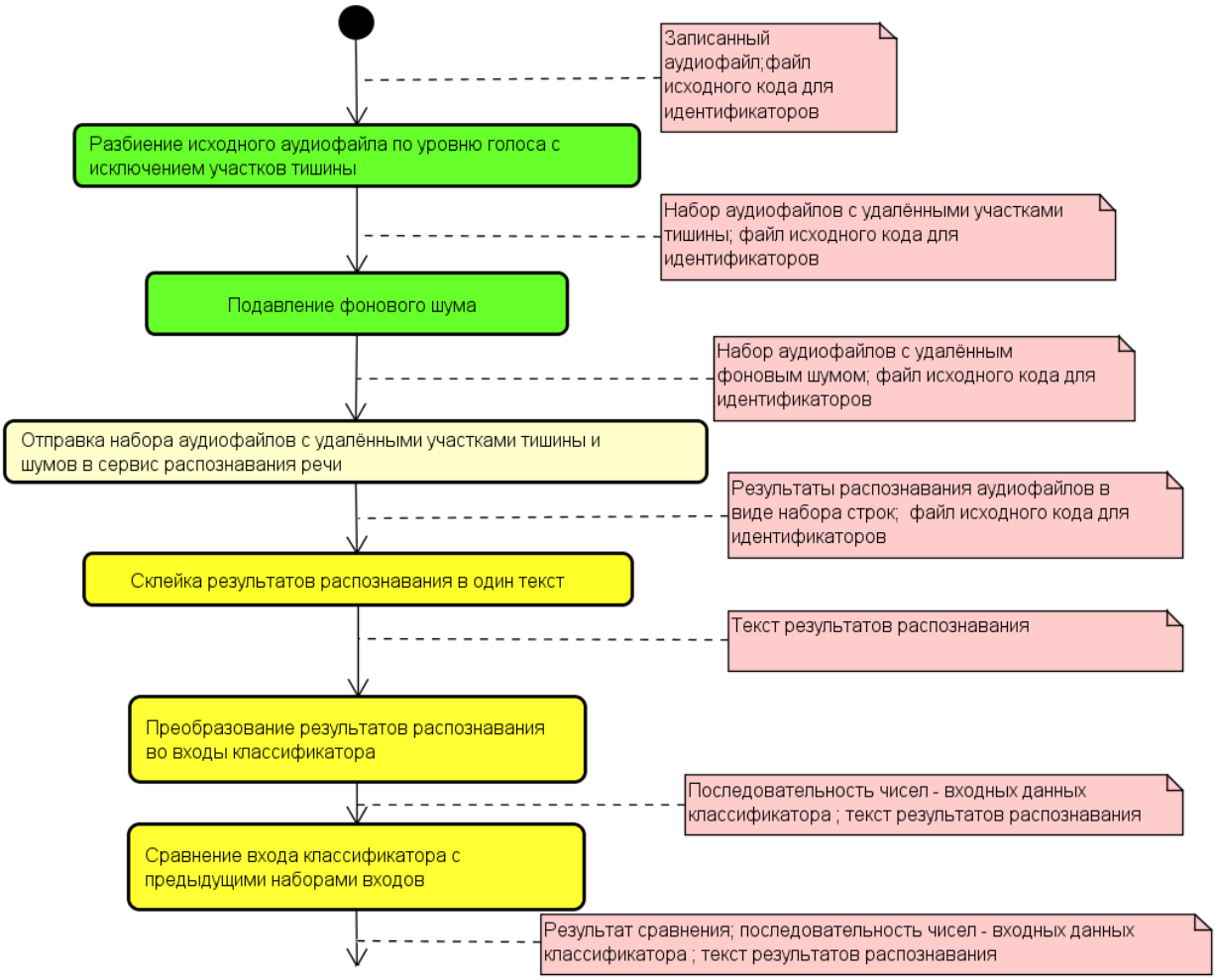

Рис. 1. – Начало работы предлагаемого подхода

Данная часть хода работы аналогична подходу, рассмотренному в предыдущей статье [2]. Однако здесь добавлены шаги преобразования результатов распознавания во входы классификатора, который представляет собой итеративное разбиение вводных данных в фонеме на основе библиотеки NLTK языка Python и преобразование на основе таблиц символов и фонем в набор чисел. Сравнение входа классификатора с предыдущими наборами входов осуществляется на основе редакционных расстояний и порогового значения. Стоит отметить, что результирующий набор

вычислений считается похожими на предыдущие результаты, если результирующее расстояние оказывается ниже данного порогового значения.

Дальнейший ход работы нового метода показан на UML-диаграмме активности на рис. 2.

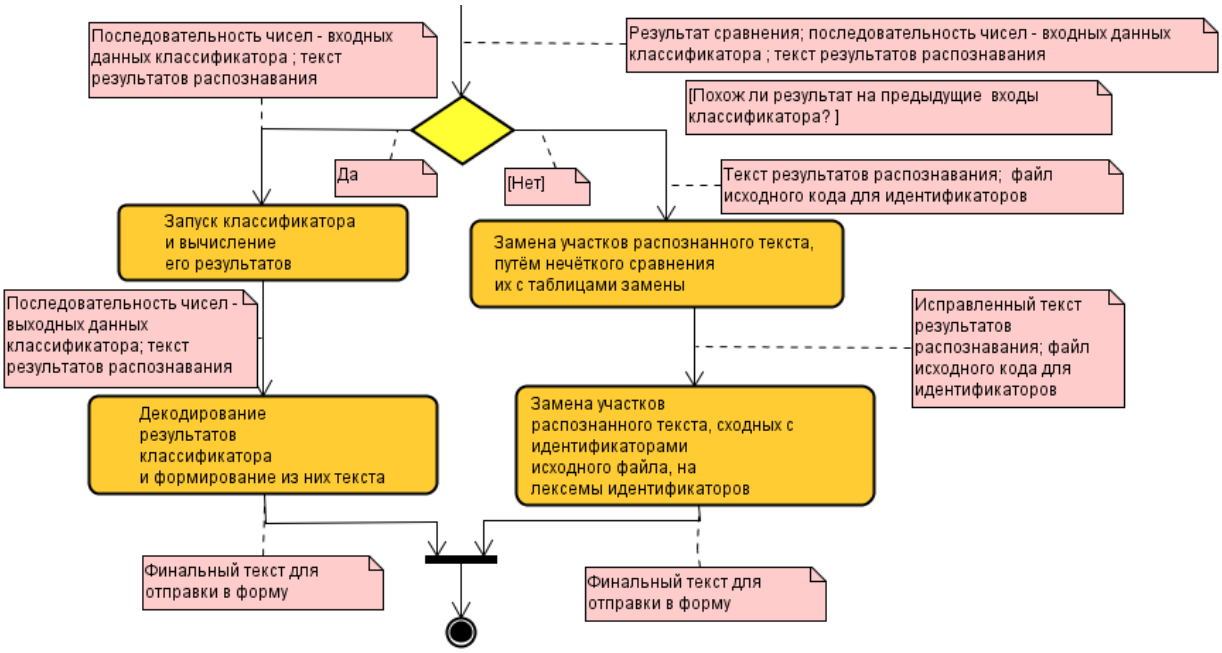

Рис. 2. – Дальнейший ход работы предлагаемого подхода

Большую часть работы на втором этапе подхода выполняет нейронная сеть, которая натренирована на записанных ранее прецедентах. Для декодирования результатов используется табличный метод, при этом для удобства выходы нейросети-классификатора могут включать как данные из таблиц, которые при декодировании заменяются на ключевые слова целевого языка, так и ссылки на идентификаторы из входного текста, что обеспечивает достаточную гибкость распознавания.

Анализ и тестирование разработанного подхода

Для тестирования подхода, разработанного на основе библиотеки PyTorch языка Python, была создана нейронная сеть-классификатор, которая была обучена на 5000 прецедентов до точности распознавания 90%. Такая точность является достаточной, так как в случае прецедентов, где нейросеть

совершала ошибку, входные данные в точности совпадали с другими прецедентами, и, соответственно имели место либо проблема распознавания со стороны сервиса Google Cloud Speech, либо аппаратные проблемы при работе микрофона, с которого велась запись.

Для тестирования был составлен список правок, включавший в себя 3 различные правки по удалению строк, 40 правок по дополнению строк, 60 правок по замене строк новыми данными. Затем было собрано 10 участников экспериментов, которые оценили свои навыки владения клавиатурой для мобильных устройств. После этого, каждый проходил замер времени, затраченного на редактирование строк в трёх различных вариантах. Первый вариант включал в себя редактирование на сайте GitHub с использованием мобильного веб-браузера. Второй вариант заключался в выполнении такой же правки в приложении Reviewgram без использования средств распознавания речи, а третий – соответственно, с применением средств распознавания речи. При этом для каждой правки было оценено количество лексем, затрагиваемых правкой, а также входящих в таблицы замены, и, наконец, количество лексем, которые входят в таблицы замены и имеют длину более 1 символа, чтобы отсечь заведомо неэффективные правки.

Анализ показал довольно серьёзное снижение времени (10–20%) на внесение правок при добавлении или замене набора длиной от 5–10 лексем (минимальная единица целевого языка, имеющая смысл) и более 10 для пользователей, неопытных в использовании мобильной клавиатуры. Для более опытных пользователей снижение времени так же присутствовало, однако для опытных в использовании мобильной клавиатуры пользователей наблюдалось, наоборот, снижение эффективности применения распознавания речи, так как набор текста вручную ими производился быстрее, чем система производила распознавание речи.

Заключение

Использование данного подхода позволяет улучшить результаты устранения ошибок распознавания речи при редактировании исходного кода программ на языках программирования высокого уровня, и, соответственно, упростить внесение правок в текст программ с мобильных устройств как для неопытных, так и для средних по опыту использования мобильной клавиатуры пользователей. Вместе с тем, однако, обучение нейросети при дальнейшем развитии может отнимать довольно много времени и вычислительных ресурсов. Для исправления данной проблемы авторами планируется дальнейшее совершенствование разработанного подхода и приложения и, возможно, изменение используемой модели нейронных сетей и подхода к их обучению.

Литература

1. Стребков Д.О., Шевчук А.В., Спирина М.О. Самостоятельная занятость на рынке удалённой работы: распространение инновационной трудовой практики // Мониторинг общественного мнения: экономические и социальные перемены. № 6. 2016. – с. 89–106.

2. Орлова Ю.А., Дмитриев А. С., Колчева Д. В. Адаптация модели распознавания речи Google Cloud Speech для упрощения редактирования исходного кода программ для ЭВМ с мобильных устройств // Инженерный вестник Дона, 2021, №2. URL: ivdon.ru/ru/magazine/archive/n2y2021/6822 (дата обращения: 12.05.2021).

3. Lewis K., Pettey M., Shneiderman B. Speech-Activated versus Mouse-Activated Commands for Word Processing Applications: An Empirical Evaluation// Int. J. Man-Machine Studies, №39. – 1993. – pp. 667–687.

4. Компаниец В.С., Лызь А.Е. Эргодизайн пользовательского интерфейса: методы юзабилити исследований // Инженерный вестник Дона, 2017, №3. URL: ivdon.ru/ru/magazine/archive/n3y2017/4333 (дата обращения: 12.05.2021).

5. Kim J., Lu C., Calvo R., McCabe K. L. A Comparison of Online Automatic Speech Recognition Systems and the Nonverbal Responses to Unintelligible. 2019. 13 p.

6. Glasser A. Automatic Speech Recognition Services: Deaf and Hard-of-Hearing Usability // Extended Abstracts of the 2019 CHI Conference. – 2019. – pp. 1-6.

7. Keane M. T., Kenny E. M. How Case Based Reasoning Explained Neural Networks: An XAI Survey of Post-Hoc Explanation-by-Example in ANN-CBR Twins // Proceedings of the 27th International Conference on Case Based Reasoning (ICCBR-19). – 2019. – URL: arxiv.org/ftp/arxiv/papers/1905/1905.07186.pdf (дата обращения: 12.05.2021).

8. Лила В.Б. Алгоритм и программная реализация адаптивного метода обучения искусственных нейронных сетей // Инженерный вестник Дона, 2012, №1. URL: ivdon.ru/magazine/archive/n1y2012/626 (дата обращения: 12.05.2021).

9. Sutskever I.,Vinyals, Le Q. V. Sequence-to-Sequence Learning with Neural Networks // Proceedings of the 27th International Conference on Neural Information Processing Systems – 2014. – URL: arxiv.org/pdf/1409.3215.pdf (дата обращения: 12.05.2021).

10. Kingma D. P., Welling M. An Introduction to Variational Autoencoders // Foundations and Trends in Machine Learning, т. 12. – 2019. – URL: arxiv.org/pdf/1906.02691.pdf (дата обращения: 12.05.2021).

References

1. Strebkov D. O., Shevchuk A. V., Spirina M.O. Monitoring obshchestvennogo mneniya: Ekonomicheskie i social'nye peremeny. № 6. 2016. pp. 89-106.

2. Orlova, Y.A., Dmitriev A.S., Kolcheva D.V. Inzhenernyj vestnik Dona, 2021, №2. URL: ivdon.ru/ru/magazine/archive/n2y2021/6822 (accessed 12.05.2021).

3. Lewis K., Pettey M., Shneiderman B. Int. J. Man-Machine Studies, №39. 1993. pp. 667-687.

4. Kompaniets V.S., Lyz' A.E. Inzhenernyj vestnik Dona, 2017, №3. URL: ivdon.ru/ru/magazine/archive/n3y2017/4333 (accessed: 12.05.2021).

5. Kim J., Lu C., Calvo R., McCabe K. L. A Comparison of Online Automatic Speech Recognition Systems and the Nonverbal Responses to Unintelligible Speech. 2019. 13 p.

6. Glasser A. Extended Abstracts of the 2019 CHI Conference. 2019. pp. 1-6.

7. Keane M.T., Kenny E. M. Proceedings of the 27th International Conference on Case Based Reasoning (ICCBR-19). URL: arxiv.org/ftp/arxiv/papers/1905/1905.07186.pdf (accessed: 12.05.2021).

8. Lila V.B. Inzhenernyj vestnik Dona, 2012, №1. URL: ivdon.ru/magazine/archive/n1y2012/626/ (accessed: 12.05.2021).

9. Sutskever I.,Vinyals, Le Q. V. Proceedings of the 27th International Conference on Neural Information Processing Systems, 2014. URL: arxiv.org/pdf/1409.3215.pdf (accessed: 12.05.2021).

10. Kingma D. P., Welling M. Foundations and Trends in Machine Learning, vol. 12, 2019. URL: arxiv.org/pdf/1906.02691.pdf (accessed: 12.05.2021).