УДК 004.522

НЕКОТОРЫЕ ПРОБЛЕМЫ УСТРОЙСТВ ПРИ УПРАВЛЕНИИ ИХ ГОЛОСОМ

Пляхина Д.В., Шкурко А.А.
Научный руководитель – старший преподаватель Михальцевич Г.А.

Еще 10 лет назад автоматическое распознавание речи находилось на грани возможного, чтобы стать главным средством взаимодействия людей с их основными вычислительными устройствами.

В преддверии эры голосовой электроники исследователи Массачусетского технологического института создали маломощный чип, специализированный для автоматического распознавания речи. В то время как мобильный телефон, работающий с программным обеспечением распознавания речи, может потребовать около 1 Вт мощности, новый чип требует от 0,2 до 10 мВт, в зависимости от количества слов, которые он должен распознать.

В реальном приложении это, вероятно, приводит к экономии энергии от 90 до 99%, что может сделать голосовое управление практичным для относительно простых электронных устройств. Это включает в себя устройства с ограниченным энергопотреблением, которые должны собирать энергию из окружающей среды или работать месяцы между зарядами батарей. Такие устройства образуют технологическую основу того, что называется “интернетом вещей”, который относится к идее, что транспортные средства, бытовая техника, строительные конструкции, производственное оборудование и даже домашний скот скоро будут иметь датчики, которые передают информацию непосредственно на сетевые серверы, помогая в обслуживании и координации задач.

"Речевой ввод станет естественным интерфейсом для многих носимых приложений и интеллектуальных устройств”, - говорит Ананта Чандракасан, профессор электротехники и компьютерных наук Ванневара Буша в Массачусетском технологическом институте, чья группа разработала новый чип. "Миниатюризация этих устройств потребует иного интерфейса, чем сенсорный или клавиатурный. Крайне важно внедрить речевую функциональность локально, чтобы сэкономить системное энергопотребление по сравнению с выполнением этой операции в облаке".

”Я не думаю, что мы действительно разработали эту технологию для конкретного приложения", - добавляет Майкл Прайс, который руководил разработкой чипа в качестве аспиранта Массачусетского технологического института по электротехнике и информатике и теперь работает в компании Chipmaker Analog Devices. “Мы попытались создать инфраструктуру, чтобы обеспечить лучшие компромиссы для системного дизайнера, чем они имели бы с предыдущей технологией, будь то программное или аппаратное ускорение”.

Прайс, Чандракасан и Джим Гласс, старший научный сотрудник Лаборатории компьютерных наук и искусственного интеллекта Массачусетского технологического института, описали новый чип в статье

Прайса, представленной недавно на международной конференции по твердотельным схемам.

Сегодня наиболее эффективные распознаватели речи, как и многие другие современные системы искусственного интеллекта, основаны на нейронных сетях, виртуальных сетях простых информационных процессоров, грубо смоделированных на работе человеческого мозга. Большая часть схем нового чипа связана с максимально эффективным внедрением сетей распознавания речи.

Но даже самая энергоэффективная система распознавания речи быстро разрядит батарею устройства, если оно будет работать без перерыва. Таким образом, чип также включает в себя более простую схему “обнаружения голосовой активности”, которая отслеживает окружающий шум, чтобы определить, может ли это быть речь. Если ответ положительный, то чип запускает большую, более сложную схему распознавания речи.

На самом деле, для экспериментальных целей, чип исследователей имел три различных схемы обнаружения голосовой активности, с разной степенью сложности и, следовательно, различными требованиями к потребляемой мощности. Какая схема составления алгоритмов является наиболее энергоэффективной, зависит от контекста. В тестах, имитирующих широкий диапазон условий, наиболее интеллектуальная программа из нескольких вариантов, привела к наибольшей экономии энергии для системы в целом. Даже при том, что он потреблял почти в три раза больше энергии, чем простейшая схема, он генерировал гораздо меньше ложных срабатываний; более простые схемы часто пережевывали свою экономию энергии, спонтанно активируя остальную часть чипа.

Типичная нейронная сеть состоит из тысяч обрабатывающих "узлов", способных выполнять только простые вычисления, но плотно связанных друг с другом. В Сети того типа, который обычно используется для распознавания голоса, узлы располагаются слоями. Голосовые данные поступают в нижний слой сети, узлы которого обрабатывают и передают их узлам следующего слоя, узлы которого обрабатывают и передают их следующему слою и так далее. Вывод верхнего слоя указывает на вероятность того, что голосовые данные представляют собой определенный речевой звук.

Сеть распознавания голоса слишком велика, чтобы поместиться в встроенной памяти чипа, что является проблемой, потому что выход за пределы чипа для получения данных является гораздо более энергоёмким, чем извлечение его из локальных хранилищ. Таким образом, программное обеспечение исследователей MIT концентрируется на минимизации объема данных, которые чип должен извлекать, из внешних запоминающих устройств.

Управление полосой пропускания

Узел в середине нейронной сети может принимать данные от дюжины других узлов и передавать их еще дюжине. Каждое из этих двух десятков соединений имеет связанный с ним “вес” - число, указывающее, насколько данные, передаваемые через него, должны учитываться в вычислениях принимающего узла. Первый шаг к минимизации пропускной способности

памяти нового чипа – это сжатие объёма памяти, связанных с каждым блоком обработки данных. Данные распаковываются только после того, как они введены на чип.

Чип также использует тот факт, что при распознавании речи волна за волной данные должны проходить через сеть. Входящий звуковой сигнал разбивается на 10-мс интервалы, каждый из которых должен оцениваться отдельно. Чип исследователей Массачусетского технологического института одновременно вводит один узел нейронной сети, но он передает через него данные с 32 последовательных 10-мс интервалов.

Если узел имеет дюжину выходов, то 32 прохода обработки данных приводят к 384 выходным значениям, которые чип хранит локально. Каждый из них должен быть соединен с 11 другими значениями при подаче на следующий уровень обработки и так далее. Таким образом, чип в конечном итоге требует значительного количества памяти для своих промежуточных вычислений. Но он извлекает только один сжатый блок памяти из внешних устройств памяти за один раз, сохраняя своё быстродействие при низком потреблении энергии.

“Для следующего поколения мобильных и носимых устройств, крайне важно обеспечить распознавание речи при сверхнизком энергопотреблении", - говорит Мариан Верхельст, профессор микроэлектроники Католического университета Левена в Бельгии. "Это происходит потому, что существует явная тенденция к устройствам с минимальными размерами, таким как часы, наушники или очки, требующим пользовательского интерфейса, который может обойтись без сенсорного экрана. Речь управления представляет очень естественный способ взаимодействия с такими устройствами”.

Хотя технология распознавания речи значительно улучшилась за последние годы, голосовые пользовательские интерфейсы все еще страдают от ошибок синтаксического анализа или транскрипции, при которых речь пользователя интерпретируется неправильно. Эти ошибки, как правило, особенно распространены, когда в речевом содержании используется технический словарь (например, медицинская терминология) или нетрадиционные варианты написания, такие как музыка или название песен. Поэтому эффективное системное проектирование для максимального понимания разговорной речи остается открытой областью исследований. Речевые пользовательские интерфейсы, которые интерпретируют и управляют разговорным состоянием, сложно спроектировать из-за присущей им сложности интеграции сложных задач обработки естественного языка, таких как распознавание именованных сущностей, поиск информации управления диалогами. Большинство голосовых помощников сегодня способны выполнять отдельные команды очень хорошо, но ограничены в своей способности распознавать диалог речи за пределами поставленной узкой задачи управления. Также могут возникнуть сложности в распознавании речи при осуществлении нескольких поворотов темы в разговоре.

Литература

1. Hardesty, L. Voice control everywhere / L. Hardesty // MIT News [Electronic resource]. – Mode of access: https://news.mit.edu/2017/low-power-chip-speech-recognition-electronics-0213 – Date of access: 23.11.2020.

2. Голосовой интерфейс пользователя [Электронный ресурс]. – Режим доступа: https://ru.qaz.wiki/wiki/Voice_user_interface – Дата доступа: 22.11.2020.