



НАУЧНЫЙ ЖУРНАЛ НАУКА И МИРОВОЗЗРЕНИЕ

СЛОЖНОСТИ РАЗВИТИЯ ТЕХНОЛОГИЙ РАСПОЗНАВАНИЯ РЕЧИ

Локтионов Никита Павлович

магистрант 1 курса, кафедра «Информационные системы и радиотехника», ИСОиП (Ф) ДГТУ, РФ, г. Шахты

Ляшов Максим Васильевич

научный руководитель, кандидат технических наук, доцент кафедры «Информационные системы и радиотехника», ИСОиП (ф) ДГТУ, РФ, г. Шахты

Одно из первых устройств для распознавания речи было создано в 1952 году. Такое устройство определяло цифры, произнесенные человеком. Через двенадцать лет, в 1964 году была показана система IBM Shoebox, которая распознавала цифры от нуля и до девяти и около шестнадцати слов [2]. Бурный рост развития систем и технологий распознавания речи пришелся на девяностые годы. В это время начали появляться программы речевого распознавания и синтеза речи коммерческого назначения. Следует отметить, что качество большей части программных решений было на очень низком уровне.

В настоящее время ведутся дорогостоящие исследования для решения проблем понимания и распознавания речи. Цель подобных исследований – разработать технологию, на основе которой можно будет создать достаточно надежную и эффективную систему распознавания речи с гарантированно высоким уровнем достоверности в режиме реального времени. Если рассматривать применение технологии с практической точки зрения, то очевидным становится формирование требований и условий, в которых будет работать автоматическая система распознавания [6]. Таких условий может быть огромное множество, но в действительности самые неприятные сложности возникают при следующих условиях:

- Произвольность и непредсказуемость пользователя системы, невозможность предугадать потенциальные действия пользователя;
- Спонтанная и сбивчивая речь, которая сопровождается ошибками и словами-паразитами; различные патологии и нарушения речи;
- Акустические искажения и помехи – как статические, так и динамические;
- Разнообразные речевые помехи.

При проведении детального анализа существующих на сегодняшний день технологий несложно заметить, что для создания достоверной системы распознавания речи эти технологии не подойдут (это касается функции помехозащищенности и условий нормализации входного сигнала), так как большинство современных систем распознавания строятся на основе скрытых марковских моделей, а эти статистические модели основываются на случайных процессах.

На сегодняшний день основное развитие систем распознавания – преобразование речевой информации в текстовую [3]. Производители анонсируют системы, которые имеют

возможность распознавать сотни и тысячи слов с точностью, достигающей до 98%. Однако подобные заявления не более чем маркетинговый ход, так как механизмы и аппараты проверки довольно слабы и неясны. В связи с этим необходимо подробно пояснить все возможные недостатки современных автоматических систем речевого распознавания:

- При первоначальной инициализации и конфигурации системы она должна настроиться на голос пользователя. Пользователь должен наговаривать текст системе несколько минут, а то и часов;
- Верификация выдает результат в районе 5% ошибок;
- Проблема омонимии. Количество словесных ошибок даже в аккуратно и четко произнесенном тексте может возрастать в несколько раз. В связи со слабой отладкой механизма коррекции ошибок быстрое отслеживание подобных ошибок практически невозможно (исключения составляют разве что совсем нелепые тексты);
- Настоящая эффективность систем распознавания неизвестна, однако, упоминается, что эффективность распознавания слов редко превышает одну треть даже для корректно подготовленных произнесенных текстов;
- Довольно длительное время обработки речи (в пределах минуты).

Все эти недостатки наглядно показывают, что системы «речь – текст» будут интересны в практическом применении лишь при полном понимании работы модели, которая будет использоваться в этих системах, а также при адекватной работе аппарата проверки их практической применимости и перспективности [6].

Основная модель, используемая почти во всех системах распознавания – скрытая марковская модель. Это вероятностная модель, предложенная Марковым А.А. в 1913 году, поначалу использовалась для аналитического разбора письменных текстов. В 70-х годах XX века эту модель успешно адаптировали для создания технологии автоматического распознавания речи [2]. При использовании этого метода возникают два вопроса, – на каких характеристиках будет основываться анализ, и какое количество скрытых сегментов в передаваемом сообщении. С первым вопросом все довольно просто, так как результат анализа, в основном, зависит от корректного сбора и обработки полученных статистических данных. Признаки, которые описывают речевые сигналы (уровни в спектре, кепстральные коэффициенты, линейный прогноз) используются в свободной последовательности и не имеют явного преимущества друг перед другом. Второй вопрос решается аналогичным образом. Выяснилось, что сегменты речи (фонемы, аллофоны и т.д.) имеют одинаковую информационную ценность для распознавания фраз, слов и других речевых единиц. Это означает, что у сегментов марковской модели почти отсутствует лингвистическая привязка и эти сегменты имеют вероятностный смысл. Из этого можно вывести неочевидное предположение, что скрытая марковская модель предполагает организацию речевого поведения, основанную на вероятности.

Также стоит сказать несколько слов о помехозащищенности. Первичная задача этого фактора – сохранение смысла передаваемой информации при акустических искажениях и помехах, а также ситуационных нарушений процессов образования и восприятия речи. Помехозащищенность может обеспечиваться по-разному. Для создания надежной помехозащищенности используются принципы дублирования и параллельной работы системы признаков, предназначенных для выделения элементов входного речевого сигнала. Проблема заключается в том, что эти системы признаков будут определяться

различными контекстами (помеховый, прагматический, семантический) в зависимости от конкретной ситуации. Такой подход сразу исключает применение вероятностной модели, однако, вероятностные методы можно задействовать на низшем уровне распознавания.

Подытоживая вышесказанное, можно выделить три главные проблемы концепции распознавания речи. Первая проблема – как будет выглядеть модель распознавания речи, если не использовать популярную, но малопродуктивную вероятностную модель? Тут многое зависит от поставленных задач и применения системы. Перспективной разработкой в этой области является стохастическая технология отечественного ученого Насыпного В.В. [7]. Вторая проблема – на чем должен основываться выбор описания первичного речевого сигнала? Эта проблема тесно граничит с возможностями понимания человеком процессов речеобразования. Здесь большие надежды подает квантовая теория речи, но пока что в большинстве систем используется стандартный статистический анализ акустических параметров речи [5]. Также перспективным выглядит использование нейронных сетей [1]. Третья проблема – проблема взаимодействия первичных речевых признаков с более высокими уровнями (семантика, прагматика) при отказе от классической линейной модели входного речевого сигнала. Здесь основными задачами является преодоление проблем омонимии, «словесного мусора», помех различного типа, а также корректного построения базы знаний, системы предварительного обучения и анализа.

Для решения небольших задач в области распознавания речи можно применять любую (даже самую простую) модель, которая может быть применима в данной области. Но для глобальных задач речевого взаимодействия человека и компьютера необходимо достаточно четко представлять механизмы работы речевой системы в целом.