



КОМПЬЮТЕРНЫЕ СИСТЕМЫ МАССОВОГО ОБСЛУЖИВАНИЯ И РЕЧЕВЫЕ ТЕХНОЛОГИИ¹

В. А. Жожикашвили, Н. В. Петухова, М. П. Фархадов

Институт проблем управления им. В. А. Трапезникова, г. Москва

Исследованы современное состояние и тенденции развития компьютерных систем массового обслуживания и речевых технологий. Отмечено, что несмотря на научные достижения в области речевых технологий, компьютерные распознаватели речи еще не достигли уровня понимания свободной речи человеком. Показано, что при нынешнем уровне распознавания можно достигнуть устойчивой и эффективной работы системы с распознаванием речи путем настройки интерфейса и управления диалогом. Изложены подходы к повышению качества систем обслуживания с речевыми технологиями алгоритмическими методами.

ВВЕДЕНИЕ

Компьютерные системы массового обслуживания появились на Западе в середине прошлого века. Первыми, кто стал их применять в своей деятельности, были крупные авиакомпании Америки и Европы. Система SABRE, созданная для авиакомпании “American Airlines”, была предназначена для автоматизации бронирования мест и управления ресурсами авиакомпании и стала первой крупной компьютерной системой для обслуживания населения.

В СССР создание первой системы такого типа было начато в 1960-х гг. в Институте проблем управления (ИПУ) в виде крупномасштабного проекта, получившего название “Автоматизированные системы массового обслуживания” (АСМО). В дальнейшем ИПУ был определен как головная организация проекта. Первая конкретная АСМО была построена для продажи билетов на самолеты и названа системой “Сирена”.

Отличительная черта АСМО состоит в наличии центральной базы данных, контролирующей быстромесяющиеся ограниченные ресурсы и информацию, и большого числа удаленных терминалов, обслуживающих население путем обращения к этой базе с помощью коротких транзакций.

Развитие АСМО шло по пути наращивания числа терминалов и внедрения сетевых технологий, обеспечивающих взаимодействие между собой теперь уже самих центров обработки данных и расширяющих возможности доступа терминалов к “чужим” ресурсам. Здесь вновь

лидерами были авиакомпании, использовавшие для организации работы по новым принципам специализированную сеть SITA. Тогда же появилось разделение центров обработки на распределительные и инвенторные. В СССР также была создана сеть “Сирена”, одна из передовых в мире по своим масштабам и принципам работы. В ней уже в то время (1970—1980 гг.) была применена пакетная коммутация, адаптивная маршрутизация, электронная почта.

Появление Интернета было встречено разработчиками этих систем с большим интересом. Они стали одними из первых, кто предоставил доступ к своим ресурсам с компьютеров пользователей. Появление нового класса пользователей сделало системы более открытыми, но и создало ряд новых проблем, связанных с защитой данных и с управлением ресурсами, ставшими более доступными.

В эти же годы наблюдается стремительный рост числа автоматизированных систем массового обслуживания и расширение областей их применения. Сейчас уже трудно представить себе компанию, научное учреждение или правительственную организацию, которая не предоставляла бы информацию о себе в Интернете, или сервисные службы, торговые организации, транспортные предприятия, не предлагающие справки или бронирование своих ресурсов пользователям Интернета.

В последние годы системы массового обслуживания расширяют круг пользователей и увеличивают доступность своих ресурсов с помощью новых технологических достижений. Речь идет, в частности, о речевых технологиях (компьютерном распознавании речи, синтезе речи, идентификации голоса) и их применении в системах, предоставляющих населению различную информацию и услуги.

¹ Работа выполнена при поддержке РФФИ (проект № 05-08-18075-а).

1. РАСПОЗНАВАНИЕ И СИНТЕЗ РЕЧИ И ИДЕНТИФИКАЦИЯ ГОЛОСА

Практическому применению распознавания речи предшествовало несколько десятилетий теоретических изысканий, начало которых относится к 1950-м гг. В 1980-е гг. удалось перейти от распознавания отдельных слов к распознаванию слитной речи. В основу этих работ были положены идеи применения методов динамического программирования и распознавания образов, выдвинутые еще в 1960—1970-х гг. различными исследователями на Западе [1—7] и в Советском Союзе (работы Т. К. Винцюка, В. М. Величко, Н. Г. Загоруйко и др. [8—10]). Ключевым стал подход, предложенный в 1980-е годы и приведший впоследствии к наиболее значимым результатам, основанный на применении модели скрытых марковских цепей для вычисления исходной гипотезы по известному конечному результату [11—13]. Тогда же началось практическое применение предложенного ранее метода описания речевого сигнала с помощью характеристических векторов. Была предложена также идея использовать статистические языковые модели, характеризующие вероятность появления определенного языкового элемента в речевом сигнале.

В 1990-х гг. удалось распространить на распознавание речи некоторые новые достижения в области распознавания образов, в результате чего появились дикторнезависимые распознаватели, которые обеспечивали уже приемлемую для разработки приложений точность распознавания. Ведущими компаниями на рынке производителей систем распознавания речи стали компании “Nuance Communication”, “Philips”, “SpeechWorks” [14] и некоторые другие. Исследовательские работы проводились в Массачусетском технологическом институте [15] и университетах США, Японии и Европы. Большое внимание и финансовую поддержку разработкам в области распознавания речи в течение многих лет оказывало DARPA (Defence Advanced Research Projects Agency). Появились первые внедрения: получение информации об авиарейсах, справки о состоянии банковского счета и услугах банка, о биржевых котировках, о погоде.

Двухтысячные годы характеризуются ростом числа внедрений, комплексным применением в приложениях распознавания и синтеза речи, а также верификации клиентов по голосу. Разрабатываются инструменты для облегчения создания приложений. Научные разработки направлены на улучшение качества распознавания, распознавание естественной речи, понимание речи, исследование мультимодальной коммуникации (совместный анализ визуальной и аудиоинформации).

Первые работы по распознаванию диктора появились в 1960-х гг., т. е. несколько позже, чем работы по распознаванию речи [16—18]. Разрабатывались как текстозависимые методы идентификации [19, 20], так и методы идентификации личности по произвольной слитной речи [21, 22]. Было предложено много идей преодоления сложностей, возникающих из-за неустойчивости речевого сигнала, влияния шумов, эмоционального состояния диктора, состояния его речевого аппарата. Интерес к идентификации голоса объясняется высокой потребностью в системах аутентификации в раз-

ных областях жизнедеятельности человека, поэтому много внимания уделялось также защите от несанкционированного доступа. Например, был предложен метод, при котором ключевые предложения, предлагаемые для произнесения, меняются каждый раз, когда система аутентификации используется очередным зарегистрированным пользователем. Поскольку эти ключевые фразы выбираются системой из своего словаря случайным образом, человек не знает заранее, что он должен будет произнести.

В работах по синтезу речи достигнуты довольно хорошие практические результаты, и для многих языков созданы “искусственные дикторы”, которые уже успешно работают на практике и вполне удовлетворительно имитируют произвольную слитную речь человека. Созданы такие продукты и для русского языка [23, 24].

В России практическое применение распознавания речи первоначально было связано с работами ИПУ РАН [25—27]. В этих работах компьютерное распознавание речи рассматривалось в блоке с АСМО, что позволяло модернизировать существующие системы и создавать новые таким образом, что клиенты могли получать справочную информацию и другие услуги, такие как заказ билетов на виды транспорта или зрелищные мероприятия, заказ такси, самостоятельно в режиме самообслуживания, без участия операторов. В 2004 г. распознавание речи, использующее речевые блоки, созданные в ИПУ, было внедрено в автоматический сервис “Телепат” электронной платежной системы “Web-Money Transfer” [28], благодаря чему ее клиентам была предоставлена возможность регистрации в системе и выполнения операций со своими счетами, используя голос, а не нажимая кнопки телефона, как раньше. Другие фирмы начали разрабатывать системы с распознаванием речи для контакт-центров с целью обеспечения возможности получения рутинных справок без участия операторов.

Практика показывает, что системы массового обслуживания с распознаванием речи могут быть весьма крупномасштабными, обрабатывающими тысячи транзакций в сутки и распознающими тысячи слов. Они позволяют эффективно заменить труд многих операторов (десятков и сотен человек) и достичь высокой рентабельности.

В США, где зарплата существенно больше, чем в России, системы с распознаванием речи рентабельны даже в небольших контакт-центрах, и спрос на них велик. В России процесс только начинается, но прогнозы весьма оптимистичны. Эксперты на Западе утверждают, что объем этого рынка в ближайшие годы многократно увеличится и в совокупности достигнет нескольких миллиардов долларов.

В применении компьютерного распознавания речи в первую очередь заинтересованы компании, работающие с большим числом клиентов и желающие перейти на новый уровень общения с ними. Среди них:

- всевозможные справочно-консультационные и сервисные службы;
- телекоммуникационные операторы связи, операторы сотовой связи, Интернет-провайдеры;
- банки, крупные финансовые структуры, страховые компании;



- крупные торговые предприятия, магазины “на диване”, службы доставки заказов;
- справочные службы вокзалов и аэропортов, службы заказа билетов и такси;
- крупные медицинские учреждения;
- туристические фирмы;
- предприятия, которым необходимо систематизировать обслуживание потока вызовов клиентов в едином унифицированном формате.

Здесь всюду на первый план выступают экономика и удобство клиентов.

Отдельную область образуют офисы, в которых государственная или муниципальная администрация проводит различные плановые мероприятия и где необходимо в короткие сроки обслужить большое число людей. Эту область можно назвать управлением очередями.

Наконец, можно указать на еще одну область применения компьютерного распознавания речи — на управление технологическими процессами. Сюда относятся управление объектами, опасными для жизни людей и окружающей среды, а также всевозможные ситуации, когда бывают заняты другой работой руки или глаза оператора, в том числе и управление транспортом на дорогах.

2. УСТОЙЧИВОСТЬ СИСТЕМ С РАСПОЗНАВАНИЕМ РЕЧИ

В Институте проблем управления выполнен ряд работ по распознаванию речи, в целях исследования свойств распознавателей, создания полнофункционального речевого интерфейса на русском языке для систем массового обслуживания и определения методов повышения качества и устойчивости работы таких систем.

Главная проблема в области компьютерного дикторезависимого распознавания речи по-прежнему заключается в точности распознавания. Точность, превышающая 95 %, практически не достигнута, и преодоление этой границы потребует, по-видимому, больших усилий и времени, если не возникнут новые идеи и принципы. На точность распознавания влияют не только характеристики распознавателя, но и шум и помехи в телефонных линиях, посторонние звуки, сбой при передаче речи по каналам связи, обрыв связи по разным причинам, поведение клиента при его речевом общении с автоматом.

Введем в рассмотрение следующее понятие устойчивости систем с распознаванием речи: устойчивость — это способность системы продолжать работу, несмотря на противодействие внешней среды и ошибки в распознавании.

Для повышения устойчивости систем с распознаванием речи применяются в основном так называемые алгоритмические или логические методы, куда относятся создание оптимального сценария, адаптивное управление диалогом, контроль сообщаемой информации, подбор подсказок, исключение петель и тупиков в диалоге.

Основными количественными критериями, характеризующими устойчивость, являются вероятность успешного завершения диалога пользователем с системой и относительная продолжительность диалога (по сравнению со временем, затрачиваемым при обработке запроса абонента человеком-оператором). Исходя из этих

критериев, оценивают варианты сценариев, управляющих стратегий, выбирают подсказки и т. д.

Логические методы повышения устойчивости весьма разнообразны и в большой степени зависят от приложения. Приведем несколько примеров.

Во многих приложениях в самом начале диалога вызывает целесообразно запросить у клиента его контактный телефон и зафиксировать его. Такой простой прием оказывается весьма универсальным, ибо работает во всех случаях нарушения связи или прекращения диалога со стороны клиента. С помощью контактного телефона сотрудник обслуживающей системы продолжат и завершат диалог с клиентом. Способность системы не потерять клиента очень важна, когда система функционирует в развитой рыночной среде и когда потеря клиента непосредственно приводит к потере дохода. Так обстоит дело в транспортных системах, например, при заказе билетов на самолеты или заказе такси.

Еще один логический механизм повышения устойчивости основан на применении контроля корректности распознанных данных, когда в систему заранее заложено знание возможных ошибок. Так, ни один месяц не может иметь более 31-го дня; время дня не может превышать 24 ч. Нельзя заказать такси на вчера или на 15 часов, если в момент заказа уже 16 часов, и много других случаев. Такие ошибки выявляются и корректируются программно.

Другая ситуация встречается, когда ошибка логически не проявляется, например, если исказился номер контактного телефона, оставаясь в пределах, например, семи знаков, или номер дома клиента — заказчика такси. В этих случаях целесообразно предусмотреть переспрос со стороны сервера.

Переспрос может быть неэффективным, если в диалоге встречаются “плохие” слова — с высокой вероятностью искажения типа: слово $i \rightarrow$ слово j . В этом случае можно прибегнуть к переспросу несколько раз подряд. Если бы система знала все “плохие” слова, то при искажении типа $i \rightarrow j$, которое отличается высокой вероятностью, она автоматически угадала бы слово i . Поиск истинного слова происходил бы быстрее, чем при переспросах. Для этого необходимо заложить в систему знание такого сорта переходов. Целесообразно сделать так, чтобы система в процессе эксплуатации сама собирала статистику ошибок.

Существуют и другие вероятностные методы исправления ошибок распознавания. Например, по аналогии с так называемым кодовым расстоянием в теории кодирования может быть введено в рассмотрение фонетическое расстояние между словами. Чем больше расстояние между парой $i - j$, тем меньше вероятность перехода $i \rightarrow j$ или $j \rightarrow i$ в процессе распознавания. Такой метод позволяет развить математические подходы и численно оценивать вероятность искажения.

3. ПРАКТИЧЕСКИЕ РЕЗУЛЬТАТЫ

Результаты исследований были применены в диспетчерском центре такси, предназначенной для автоматизации приема заказов такси и использующей речевой интерфейс для общения с клиентами.

Диалог начинается с приветствия и развивается далее по трем направлениям: справки о тарифах, стоимость конкретной поездки, заказ такси.

В приветствии называется имя службы и сообщается о предоставляемых услугах. Затем клиенту предлагается выбрать услугу, которой он хочет воспользоваться.

Если клиент выбирает услугу “Справки о тарифах”, то диалог развивается вокруг различных тарифов, которые может предложить служба заказа такси (дневной тариф, ночной, выходного дня, тариф “аэропорт”).

Если клиент выбирает услугу “стоимость конкретной поездки”, компьютер сообщает ему, какая информация требуется для вычисления стоимости поездки (адрес подачи такси, конечный пункт, дата и время подачи). Подготовленный таким образом клиент отвечает далее на вопросы компьютера, сообщая все эти данные. По ходу сбора данных о поездке происходит взаимодействие сервера распознавания с сервером-картой, в котором определяются координаты маршрута.

Взаимодействие осуществляется в реальном масштабе времени через механизм сокетов.

После получения всех сведений о маршруте и времени поездки происходит новое взаимодействие, на этот раз с сервером, содержащим матрицу стоимостей, и полученный ответ озвучивается для клиента.

Если клиент выбирает услугу “заказ такси”, компьютер просит, прежде всего, назвать контактный телефон для связи с клиентом водителя или диспетчера, а затем в ходе диалога просит назвать адрес подачи такси, дату и время подачи. Полученная информация через электронную почту передается диспетчеру, о чем сообщается клиенту.

После завершения обслуживания по любой из ветвей диалога компьютер предлагает клиенту снова выбрать одну из услуг. Например, после получения справки о стоимости поездки компьютер спрашивает, не хочет ли клиент заказать такси и будет ли он менять маршрут для заказа такси. Если клиента устроила названная стоимость поездки по интересующему его маршруту, и он решает заказать такси, компьютер просит назвать номер контактного телефона и передает диспетчеру заказ с полученными ранее сведениями о маршруте и времени подачи такси.

В начале диалога компьютер информирует клиента, что тот может на любом шаге диалога попросить помощь, сказав слово “помощь”. Распознав слово “помощь”, компьютер выдает подсказку, содержание которой зависит от места, в котором клиент просит помощи.

Помимо подсказок-помощи, имеется еще целый ряд промптов, которые выдаются системой в случае обнаруженных ошибок, нераспознавания или превышения тайм-аута на ожидание ответа пользователя. Назначение этих промптов состоит в том, чтобы помочь клиенту сформулировать ответ правильным образом, исправить допущенную и обнаруженную компьютером ошибку или разъяснить ему, что требуется от него на данном шаге диалога.

Компьютер может также перевести звонок на оператора. Решение о переводе звонка компьютер принимает по разным причинам: из-за повторения ошибок, в случае отсутствия у него нужной клиенту информации или по просьбе клиента.

В случае, если контактный телефон уже сообщен компьютеру, а ошибки на одном из последующих шагов диалога повторяются, компьютер посылает диспетчеру неполную заявку клиента и сообщает клиенту о том, что через несколько минут диспетчер свяжется с ним.

Для работы автоматической диспетчерской службы был разработан ряд блоков распознавания, отличающихся сложностью и объемом. Наиболее простые блоки способны распознавать до нескольких десятков слов и фраз. К ним относятся блоки для распознавания ответов типа “Да/Нет” со всеми вариантами этих высказываний, которые может употребить клиент, и блоки выбора конкретной опции из ряда предложенных.

Наибольшим по объему оказался блок для распознавания названий улиц и других объектов (вокзалы, гостиницы, аэропорты) Москвы. Он содержит почти четыре тысячи слов.

Сложной логикой отличаются блоки распознавания номеров московских телефонов, даты и времени суток. Это объясняется тем, что клиент может называть эти компоненты самым разным образом, и эти варианты произношения должны быть учтены в грамматиках.

В ходе отладки системы автоматической диспетчерской службы было проведено много испытаний и экспериментов с целью достижения устойчивой работы системы и удовлетворенности клиентов ее функционированием. В ходе этих испытаний был выявлен ряд закономерностей и зависимостей и предприняты меры для достижения поставленной цели. Основные результаты наших работ по улучшению функционирования состоят в следующем:

- показана возможность и целесообразность применения математических моделей для расчета временных характеристик речевого интерфейса;
- выбрано оптимальное число повторов в случае ошибок распознавания;
- прослежено влияние первого промпта на поведение клиентов;
- разработана стратегия управления диалогом;
- проанализировано различие в поведении новичков и опытных пользователей и предложены варианты управления диалогом в зависимости от предположения системы о типе клиента (новичок или опытный пользователь);
- запрограммирован сбор статистики в ходе функционирования системы с целью использования ее для улучшения качества функционирования.

Проведенные в соответствии с этими результатами мероприятия позволили улучшить качество функционирования системы автоматической диспетчерской службы.

ЗАКЛЮЧЕНИЕ

Научные достижения в области речевых технологий привели к созданию продуктов, пригодных к практическому применению, которое уже началось. Однако приходится констатировать, что уровня понимания свободной речи человеком компьютерные распознаватели еще не достигли. Анализ созданных систем и работа по проектированию новых показали, насколько важно применять алгоритмические методы для достижения устойчивой и эффективной работы, которая остав-



ляла бы у клиентов ощущение удобства и удовлетворенности при получении информации или услуг в режиме самообслуживания. Обеспечение этих условий позволяет уже сейчас, при достигнутом качестве работы распознавателей, создать социально значимые системы, внедрение которых сделает доступ населения к услугам и данным более дешевым, удобным и круглосуточным.

ЛИТЕРАТУРА

1. Coker M. J. An improved isolation word recognition system based upon the linear prediction residual // IEEE Trans. Acoustics, Speech, Signal Proc. — 1976. — Vol. ASSP-24. — P. 206—209.
2. Sakoe H., and Chiba S. Dynamic programming algorithm optimization for spoken word recognition // Ibid. — 1978. — Vol. ASSP-26 (1). — P. 43—49.
3. Sakoe H. Two level DP matching — a dynamic programming based pattern matching algorithm for connected word recognition // IEEE Trans. Acoustics, Speech, Signal Proc. — 1979. — Vol. ASSP-27. — P. 588—595.
4. Viterbi A. J. Error bounds for convolutional codes and an asymptotically optimal decoding algorithm // IEEE Trans. Information Theory. — 1967. — Vol. IT-13. — P. 260—269.
5. Bridle J. S., and Brown M. D. Connected word recognition using whole word templates // Proc. Inst. Acoust. Autumn Conf. — 1979. — P. 25—28.
6. Myers C. S., and Rabiner L. R. A level building dynamic time warping algorithm for connected word recognition // IEEE Trans. Acoustics, Speech, Signal Proc. — 1981. — Vol. ASSP-29. — P. 284—297.
7. Lee C. H., and Rabiner L. R. A frame synchronous network search algorithm for connected word recognition // Ibid. — 1989. — Vol. 37, N 11. — P. 1649—1658.
8. Винцюк Т. К. Распознавание слов устной речи методами динамического программирования // Кибернетика. — 1968. — № 1. — С. 81—88.
9. Величко В. М., Загоруйко Н. Г. Автоматическое распознавание ограниченного набора устных команд // Вычисл. системы. — 1969. — Вып. 36. — С. 101—110.
10. Винцюк Т. К. Анализ, распознавание и интерпретация речевых сигналов. — Наукова думка, Киев, 1987. — 262 с.
11. Ferguson J., Ed. Hidden Markov models for speech. — NJ: IDA, Princeton, 1980.
12. Rabiner L. R. A tutorial on hidden Markov models and selected applications in speech recognition // Proc. IEEE. — 1989. — Vol. 77, N 2. — P. 257—286.
13. Rabiner L. R., and Jang B. H. Fundamentals of Speech Recognition. — New Jersey: Prentice-Hall, Englewood Cliff, 1993.
14. <http://www.nuance.com>
15. <http://groups.csail.mit.edu/sls/sls-blue-noflash.shtml>
16. Pruzansky S., Mathews M. V. Talker recognition procedure based on analysis of variance // J.A.S.A. — 1964. — Vol. 40. — P. 2041—2047.
17. Li K. P., et al. Experimental studies in speaker verification using a adaptive system // Ibid. — 1966. — Vol. 36. — P. 966—978.
18. Doddington G. R. A method of speaker verification // Ibid. — 1971. — Vol. 49. 139(A).
19. Furui S. Cepstral analysis technique for automatic speaker verification // IEEE Trans. Acoustics, Speech, Signal Proc. — 1981. — Vol. ASSP-29. — P. 254—272.
20. Rosenberg A. E., Sambur M. R. New techniques for automatic speaker verification // Ibid. — 1975. — Vol. ASSP-23. — P. 169—176.
21. Li K. P., Hughes G. W. Talker differences as they appear in correlation matrices of continuous speech spectra // J.A.S.A. — 1974. — Vol. 55. — P. 833—837.
22. Beek B., et al. An assistment of the technology of automatic speech recognition for military applications // IEEE Trans. Acoustics, Speech, Signal Proc. — 1977. — Vol. ASSP-25. — P. 310—322.
23. <http://www.sakrament.com/>
24. http://www.clarosoftware.com/product_info.php/products_id/661/tab/p/cPath/219
25. Открытые системы массового обслуживания / В. А. Жожикашвили, Р. В. Билик, В. А. Вертлиб и др. // Проблемы управления. — 2003. — № 4. — С. 55—62.
26. Применение распознавания речи в автоматизированных системах массового обслуживания / В. А. Жожикашвили, А. В. Жожикашвили, Н. В. Петухова, М. П. Фархадов // Автоматизация и современные технологии. — 2003. — № 11.
27. The first voice recognition applications in Russian language for use in the interactive information systems / V. A. Zhozhikashvili, M. P. Farkhadov, N. V. Petukhova, and A. V. Zhozhikashvili // Proc. of the Ninth International Conference "Speech and Computer" SPECOM'2004. — Saint-Petersburg, 2004. — P. 304—307.

☎ (495) 334-87-10

E-mail: mairs@ipu.ru



Новая книга

Левин В.И., Костиневич В.В. Математические модели и алгоритмы планирования и оптимизации образовательных процессов. — Пенза: Изд-во Пензенской государственной технологической академии, 2005. — 128 с.

Дан обзор методов математического моделирования, планирования и оптимизации образовательных систем. Изложены новые методы моделирования и алгоритмы построения допустимых планов реализации учебного процесса на различных уровнях иерархии образовательной системы, основанные на наглядных графовых и геометрических представлениях. Описаны новые статистические алгоритмы оценки эффективности предложенных методов моделирования, планирования и оптимизации учебного процесса. Приведены сведения о компьютерных программах для планирования и оптимизации учебного процесса и анализа эффективности учебных планов. Даны общие рекомендации по внедрению разработанных планов и алгоритмов в практику.

Для студентов вузов, аспирантов, специалистов по управлению образованием, руководителей вузов и школ.