ПРОЕКТИРОВАНИЕ СИСТЕМ РЕЧЕВОГО ДИАЛОГА♦

И. В. Ли, А. Л. Ронжин

Санкт-Петербургский институт информатики и автоматизации РАН

СПИИРАН, 14-я линия ВО, д. 39, Санкт-Петербург, 199178

<{lee,ronzhin}@iias.spb.su>

УДК 004.5

Ли И. В., Ронжин А. Л. **Проектирование систем речевого диалога** // Труды СПИИРАН. Вып. 3, т. 1. — СПб.: Наука, 2006.

Аннотация. *Проектирование диалоговых систем охватывает ряд сложных проблем, таких как обработка речевого сигнала, семантический анализ, понимание смысла речи, управление диалогом, синтез речи и другие. Особое место среди них занимает проблема оптимизации управления диалогом, поскольку в свою очередь включает в себя комплекс задач, связанных с выбором модели диалога, разработкой методов управления диалогом и их обучением. В данной статье рассматриваются основные проблемы проектирования диалоговых систем, а также представлен краткий обзор современных систем речевого диалога.* — Библ. 16 назв.

UDC 004.5

Li I. V., Ronzhin A. L. **Speech dialogue systems development** // SPIIRAS Proceedings. Issue 3, vol. 1. — SPb.: Nauka, 2006.

Abstract. *The development of dialog system covers a set of different problems such as speech processing, semantic analysis, speech understanding, dialogue management, speech synthesis and others. Among them a problem of dialogue management optimization can be considered particularly since it includes the complex of tasks connected with a choice of dialogue model, development of dialogue management methods and training of them. In the framework of the paper the basic problems of dialogue systems designing are considered, and also the brief review of modern systems of speech dialogue is presented.* — Bibl. 16 items.

1. Введение

Развитие средств автоматизации передачи и обмена информацией в последние десятилетия 20 века привело к их стремительному внедрению в нашу повседневную жизнь. Далее встал вопрос повышения качества взаимодействия и уровня комфортности. Как наиболее естественное и распространенное средство человеческого взаимодействия в интеллектуальных системах стала использоваться речь. Параллельно развивались и иные технологии, такие как распознавание жестов, мимики, движений губ и другие. Постепенно стали появляться системы человеко-машинного взаимодействия, использующие различные модальности.

Современная диалоговая система должна быть эффективной, быстрой и комфортной с точки зрения человеческого восприятия. Поэтому в основе систем диалога лежат теории человеческого общения. При проектировании диалоговых систем необходимо учитывать специфику прикладной области и условия использования системы, привлекать знания экспертов предметной области и располагать инструментарием разрешения проблем непонимания между системой и человеком. Все эти вопросы напрямую связаны с разработкой конкретных методов представления, обработки и управления диалогом. В данной статье рассматриваются основные принципы проектирования диалоговых систем и проблемы оптимизации управления диалогом.

♦Данное исследование проводится в рамках Европейской научной сети SIMILAR NoE FP6-IST № 507609 и гранта ИНТАС № 04-77-7404.

Основные проблемы, связанные с разработкой системы человеко-машинного диалога: обработкой входных данных, семантическим анализом и разработкой оптимального менеджера диалога рассматриваются во 2 разделе.

В 3 и 4 разделах обсуждаются вопросы оптимизации управления диалогом: подходы к моделированию диалога и методы управления диалогом. В 4 разделе дана оценка существующих подходов к обучению стратегии диалога, рассмотрены их сильные и слабые стороны.

В 5 разделе представлены возможные области применения диалоговых систем, приводятся примеры систем речевого диалога разработанные в последние годы и получившие широкое применение.

В заключении делаются основные выводы по исследуемой теме: перечислены основные проблемы проектирования систем речевого диалога и оптимизации процесса управления диалогом.

2. Базовая архитектура диалоговой системы

В эпоху строго разделяемых научных направлений в области речевых исследований диалог рассматривался как обмен речевой информации. Другими словами, пользователь и система обмениваются речевыми сообщениями для достижения определенной цели. Однако, с развитием систем, использующих различные модальности человеко-машинного взаимодействия, диалог стал рассматриваться как многомодальный процесс [1]. Такой подход не противоречит природе человека: при взаимодействии с окружающим миром он по возможности использует все данные ему природой органы чувств (по отношению к машине это и будут различные модальности). В процессе диалога человек сопоставляет информацию, пришедшую из различных источников (органов чувств) и совмещает ее со своим представлением о мире и текущим контекстом для того, чтобы понять собеседника (машину или человека). Помимо речи для выражения своего намерения человечество научилось использовать жесты, картинки, знаки и др. При речевом взаимодействии такие способы дают возможность получить дополнительную информацию, необходимую для понимания. Таким образом, недостаток одного вида информации восполняется избыточностью информации, полученной из других источников.

Проектирование системы речевого диалога включает комплекс задач, которые необходимо решать на соответствующих уровнях обработки входной информации. Базовая архитектура типовой диалоговой системы представлена на рис 1. в виде различных подсистем объединенных потоками информации, идущими от пользователя к системе и обратно. Необходимо отметить, что потоки информации от пользователя к системе могут быть различной природы, т.е. относится к различным видам модальностей (жесты, речь, ввод с клавиатуры, движение мышью и др.). Входная информация обрабатывается и затем передается в подсистему управления диалога, которая управляет диалогом и решает, что делать дальше. В соответствии с этим решением генерируется соответствующая ответная информация, которая передается пользователю.

Поскольку человек получает информацию из внешнего мира при помощи своих ощущений, входные сигналы должны быть, в первую очередь, собраны и трансформированы соответствующими сенсорными системами. Датчики сенсоров могут быть разных видов, например, видеокамеры, массивы микрофонов, сенсорные клавиатуры, джойстики, клавиатуры, сенсорные перчатки и другие. Эти устройства обычно конвертируют аналоговые сигналы в цифровые, а затем

Труды СПИИРАН. 2006. Вып. 3. Т. 1. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2006. Issue 3. V. 1. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

передают их на обработку компьютерам или специализированным вычислительным устройствам. Подсистема, которая предназначена для получения входных сигналов, отвечает за их предварительную обработку, включая устранение шумов и извлечение признаков. При этом шум может быть нескольких типов. Во-первых, это дополнительный фоновый шум, который может возникать в процессе передачи данных от источника к приемнику, например, посторонние разговоры во время синтеза или записи речи, блики от солнечного света на дисплее или тень на лице пользователя при видеонаблюдении и т.д. Во-вторых, шумы, возникающие при передаче сигналов по каналам связи. Здесь можно отметить различные электромагнитные шумы, возникающие в цепях вычислительных машин, а также потери от дополнительного кодирования сигналов при передаче их на удаленные расстояния.

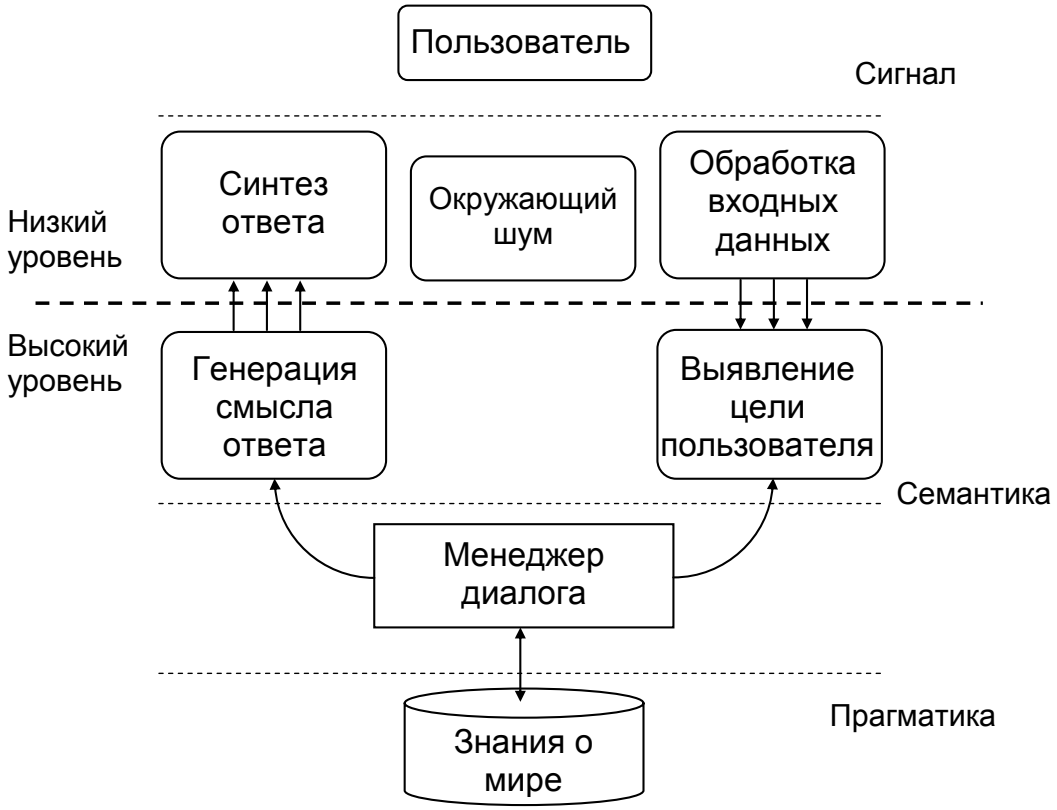

Рис. 1. Базовая архитектура диалоговой системы.

Процесс удаления шума может производиться как до, так и после процедуры параметрического представления сигнала. В случае акустического сигнала, полезно использовать методы удаления шума в исходном сигнале для того, чтобы улучшить точность извлеченных признаков. В случае распознавания жестов, шум привнесенных движений вокруг пользователя может быть подавлен сегментацией переднего и заднего планов, который применяется как к исходному сигналу, так и к полученным признакам.

Операция извлечения признаков (параметрического представления) направлена на уменьшение объема данных, которые должны быть обработаны последующими подсистемами. Действительно, например, полный видео поток является не только слишком большим объемом данных для обработки в системе диалога, но и содержит массу избыточной информации, если его нужно использовать только для распознавания жестов. Фактически, часто используются

Труды СПИИРАН. 2006. Вып. 3. Т. 1. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2006. Issue 3. V. 1. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

только некоторые характерные параметры (например, для видео это - параметры интересующей области и ее движения), которые передаются остальной системе дальше.

Характерные признаки извлекаются из входного сигнала и передаются к подсистеме смыслового анализа для «Выявления цели пользователя». Как уже было сказано ранее, мультисенсорный ввод данных может быть избыточным или дополняющим. Хотя некоторые многомодальные системы используют очень примитивные технологии слияния данных, большинство эффективных систем использует комплекс статистических моделей для получения цели и намерения пользователя. Это методы обучения, а также общие статистические методы, которые часто используются в многомодальных системах, поскольку они способны учитывать зависимости между потоками различных входных данных. Наиболее часто используются фильтры Калмана и скрытые марковские модели (СММ) являющиеся частными случаями динамических сетей Байеса, кроме того, сейчас разрабатывается концепция асинхронных СММ для совместной обработки аудио-видео сигнала [2].

Традиционные методы семантического разбора обычно опираются на идею «понимание есть способность ответить на все вопросы, связанные с входным высказыванием». Этот принцип обычно сводится к необходимости скрупулезного заполнения всех слотов семантического фрейма, что ведет к усложнению модуля управления диалогом, к переспросам и к необоснованному замедлению хода диалога. В действительности человек обычно использует интуитивное чувство ситуации и способен понимать различные сокращения, идиомы, профессиональный жаргон. Иными словами, существует некоторый разумный уровень полноты диалоговых фраз, в принципе достаточный для их надежного понимания без переспросов. Поэтому статистические данные, с достаточной полнотой отражающие речевое поведение человека, могут быть полезными в решении данной проблемы.

Теперь рассмотрим главный модуль диалоговых систем — *менеджер диалога* (МД), который отвечает за обработку как входных, так и выходных сообщений на концептуальном уровне, и отвечает за процесс коммуникации. После того как было определено намерение пользователя, система должна принять решение, что делать дальше. Модуль МД координирует взаимодействие с пользователем, а также с остальным окружением. Во-первых, система должна взаимодействовать с базами данных, например, для поиска информации, необходимой либо менеджеру для продолжения обмена с пользователем, либо самому пользователю. Во-вторых, менеджер может взаимодействовать с другими устройствами для того, чтобы собрать информацию, не касающуюся пользователя (например, температуру воздуха, время дня и т.д.), но помогающую принять решение относительно выполнения следующего действия. Всевозможные внешние устройства обеспечивают дополнительную информацию (прагматический уровень) собранную в базу знаний, которая на рис. 1 называется «Знания о мире».

После того как менеджер диалога принял решение относительно следующего действия, необходимо передать эту информацию пользователю. Этот процесс начинается с порождения набора концептов, которые выражают эту информацию. Действительно, когда от кого-то что-то требуется другим человеком в ходе диалога, он вначале думает о том, что было сказано, и затем выдает серию идей или концептов, выражающих результаты его размышлений. До того как производить какие-то действия (слова или жесты), в первую очередь, стро-

Труды СПИИРАН. 2006. Вып. 3. Т. 1. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2006. Issue 3. V. 1. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

ится концептуальная структура того, что необходимо ответить. По этому же принципу модуль генерации смысла ответа строит набор концептов, которые должны быть переданы пользователю каким-либо способом. Этот процесс иногда называют планированием. И, наконец, последний модуль в диалоговой системе предназначен для синтеза ответа с помощью речи, звуков, картинок и др., которые понятны пользователю.

Продолжая обсуждение проблем проектирования диалоговых систем, следует отметить, что реакция системы на ввод пользователя зависит от определенного пользовательского намерения, а также от специфики решаемой задачи. Поэтому менеджер диалога является еще более зависимой от задачи частью системы и, в основном, проектируется для решения определенной прикладной задачи с ограниченной предметной областью. Далее рассмотрим более детально структуру менеджера диалога.

3. Менеджер диалога

Менеджер Диалога принимает решение о том, что делать системе на основе информации, пришедшей с предыдущих модулей системы, истории диалога, информации, которую можно найти в базе «знаний о мире» и своей внутренней стратегии (отображения состояний в действия). Попробуем рассмотреть основные аспекты, связанные с построением оптимального МД, а именно, его оптимальной стратегии. Критерий оптимальности в случае диалоговых стратегий не является однозначным, и поэтому до сих пор не существует окончательного определения оптимальности, несмотря на огромное число исследований, проведенных в последнее время в области оценивания систем речевого диалога. Более того, существует несколько проблем в построении МД:

- глубина контекста, которую следует учитывать (тесно связана со способом мышления в диалоге),
- внутреннее преставление состояния,
- возможные действия, которые может выполнить МД,
- допустимый системой уровень инициативы пользователя,
- адаптация системы к окружению и пользователю,
- надежность,

все это должно быть учтено и реализовано в результате построения оптимального менеджера диалога. Поэтому далее рассматриваются проблемы построения оптимального МД.

3.1. Модели диалога

До построения автоматических систем речевого диалога были проведены исследования в области человеко-машинного диалога в целом. Основная цель состояла в разработке теории диалога, включая теорию проблемно-ориентированного диалога, в котором участники взаимодействуют для достижения некоторой определенной цели. Не все исследователи согласны с тем, что диалог человека с человеком должен служить моделью для человеко-машинных диалогов, так как люди адаптируют свое поведение во время разговора с машиной. Тем не менее, в области построения систем речевого диалога применяются четыре подхода на основе модели человеческого общения: диалоговые грамматики, сценарные модели, модели диалоговых игр и модели совместных действий.

Труды СПИИРАН. 2006. Вып. 3. Т. 1. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2006. Issue 3. V. 1. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

Диалоговые грамматики — это первый подход, разработанный для моделирования диалогов [3]. Фактически, процесс понимания естественного языка использует некоторый набор правил, накладывающих последовательные и иерархические ограничения на приемлемые диалоги, такие, как синтаксические грамматические правила, накладывающие ограничения на грамматическую допустимость высказывания. Заметим, что для построения диалоговых систем была разработана специальная (но очень похожая на теорию речевых актов) теория диалоговых актов. Самым большим недостатком диалоговых грамматик является то, что для истории текущего состояния используется только предыдущее высказывание (предыдущий речевой акт или акт диалога). Наверное, поэтому было разработано столь мало успешных систем на основе этой модели.

Сценарные модели основаны на том наблюдении, что люди выполняют действия не случайно, а планируют свои действия для достижения различных целей. В случае коммуникативных действий (речевых актов) их цели включают изменение ментального состояния слушателей. Сценарные модели диалога предполагают, что речевой акт говорящего — это часть сценария, и слушающий должен обнаружить его и ответить в соответствии с основным планом. Другими словами, при извлечении цели из высказывания система должна учитывать весь диалог и интерпретировать высказывание в контексте сценария. Эти модели оказались более сильными, чем диалоговые грамматики, однако получение цели и принятие решения в контексте сценарных диалогов иногда бывает очень затруднительным. Тем не менее, модели диалога на основе сценариев используются в системах речевого диалога.

Теория диалоговых игр — это попытка учесть идеи сценарных моделей и диалоговых грамматик в одной структуре. Предположим, что диалоги состоят из череды так называемых игр. Каждая игра составлена из последовательности ходов, которые возможны в соответствии с набором правил (похожих на грамматики), и вся игра запланирована участвующими агентами (как в модели основанной на сценариях). Таким образом, агенты совместно используют знания (представления и цели) в ходе диалога, и игры могут быть вложенными (возможны поддиалоги) для достижения подцелей. В этой структуре ходы часто приравниваются к речевым актам. Модель довольно формально определяет ходы, допустимые для каждого из участников в данный момент игры (по правилам и в соответствии с целью) и таким образом, моделируются диалоги. Эта модель доведена до реализации и используется при проектировании некоторых диалоговых систем [4].

Предыдущие подходы рассматривали диалог как результат взаимодействия генератора сценария (пользователя) и распознавателя сценария (компьютера), работающих согласованно, но это не объясняет, почему участники задают уточняющие вопросы, используют подтверждающие высказывания и т.д. Другая модель диалога рассматривает диалог как *совместную деятельность*, в которой агенты делают что-то вместе. Участие в диалоге требует от участников наличия списка совместных соглашений для понимания друг друга, и это служит причиной того, что уточнения и подтверждения столь распространены в диалогах. Это семейство моделей диалога вызывает большой интерес, и уже применялась в нескольких системах [3].

Описанные модели диалога могут использоваться в системах речевого диалога для того, чтобы объяснить семантику и построить внутреннее состояние. В соответствии с заданной моделью диалога вбираются методы управления диалогом, представленные в следующем разделе.

Труды СПИИРАН. 2006. Вып. 3. Т. 1. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2006. Issue 3. V. 1. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

3.2. Методы управления диалогом

Теперь рассмотрим методы, направленные на управление диалогом. В основном выделяют четыре метода: (1) метод доказательства теорем; (2) конечный автомат; (3) метод заполнения форм; (4) метод самоорганизации. Рассмотрим кратко каждый из этих методов.

Метод доказательства теорем был разработан для управления речевым диалогом с ограниченными параметрами [5]. В основе метода лежит идея о том, что система пытается доказать, что проблема (теорема) решена. Как и в математическом доказательстве, здесь существует несколько шагов и на каждом шаге система может использовать заведомо известные аксиомы или дедукцию. Если на данном шаге знания о мире не способны обеспечить аксиомы для управления диалогом, которые позволяют продолжить доказательство, и оно не может быть выведено из других аксиом, то аксиома рассматривается как потерянная, и система запрашивает у пользователя новую информацию. Если пользователь не может предоставить информацию, то решается новая теорема: «Пользователь может предоставить релевантную информацию». Затем производится обучение поддиалога. Этот метод был применен в Прологе как метод на основе правил, который подходит для логического программирования. Основной недостаток метода состоит в том, что стратегия является заранее фиксированной (как и шаги доказательства).

В *методах конечных автоматов* диалог представлен как сеть переходов состояний, где переходы между состояниями диалога специфицируют все возможные пути по сети. В каждом состоянии действие выбрано в соответствии со стратегией диалога, и результат действия ведет к новому переходу между состояниями.

Основным недостатком этих методов является то, что все возможные диалоги должны быть известны и описаны заранее. Тем не менее, методы переходных состояний уже широко используются в системах диалога, поскольку обеспечивают простую форму моделирования диалога, так как задача может быть напрямую отображена в структуре диалога. Метод конечного автомата легок в понимании и более нагляден для разработчика, поскольку визуальное, глобальное и эргономичное представление других методов управления отнюдь не «дружественно» пользователю.

Визуальное представление конечного автомата является гораздо более точным, поскольку диалог описан как набор состояний, связанных переходами. Модель конечного автомата не может описать системы, где диалог ведет пользователь. Однако в большинстве диалогов система является ведомой или используется стратегия со смешанной инициативой. Большинство приложений, таких как заполнение анкеты, запрос к базе данных или справочные системы, более успешно работают таким способом. Поэтому, данный метод является наиболее популярным. Даже если другие методы (самоорганизующиеся) определенно более гибки, то вложенные поддиалоги более удобно разработать с помощью конечного автомата.

Методы заполнения форм также называют фреймовыми методами. Они чаще используются в приложениях, связанных с передачей информации от пользователя к системе речевого диалога, когда информация может быть представлена как набор пар атрибут–значение. Структура атрибут–значение может рассматриваться как форма и пользователь должен определить значения каждого поля (атрибут, слот, фрейм) формы. Каждое пустое поле можно

Труды СПИИРАН. 2006. Вып. 3. Т. 1. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2006. Issue 3. V. 1. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

использовать для передачи подсказок. Стратегия диалога нацелена на полное заполнение формы и поиск значений для всех полей на основе высказываний пользователя.

В отличие от предыдущих методов управления диалогом семейство *самоорганизующегося управления* не требует предварительной спецификации всех путей диалога. Каждое действие системы и реакция пользователя требует построения новой конфигурации, которая ассоциируется с определенным поведением. Нет никакой необходимости знать, какая конфигурация имеет место. Это в основном относится к диалоговому управлению событиями. Одна из наиболее известных попыток использовать этот вид управления диалогом была предпринята Phillips в языке HDDL (Harald's Dialogue Description Language), но подход не получил широкого распространения, поскольку потребовались очень высокие затраты на его создание и развитие.

Другим важным вопросом построения стратегии диалога является о*пределение степени инициативы системы и пользователя*. Действительно, кажется очевидным, что управление диалогом будет более простым, если система всецело управляет ходом диалога до тех пор, пока пользователь удовлетворен, и если требуется, то он может взять инициативу на себя. Можно выделить три вида инициативы системы:

- активный режим: только система может задавать последовательности определенных вопросов пользователю. Пользователь может ответить на эти вопросы и предоставить только запрашиваемую информацию;
- пассивный режим: инициативу проявляет только пользователь, и он запрашивает информацию у системы. Система может корректно интерпретировать пользовательский запрос и ответить на определенные вопросы без особых уточнений;
- смешанный режим: пользователь и система совместно управляют диалогом, чтобы достичь цели пользователя. Пользователь может предоставлять дополнительную информацию, которую у него еще не запросили или просить систему выполнять определенные действия до тех пор, пока система способна управлять состояниями диалога, не отклоняясь от правильного пути диалога. Система может также брать управление на себя, поскольку в некоторых случаях (из-за шума или по другой причине) модули обработки входных сигналов могут давать сбои или недостаточно точную информацию.

Подсознательно можно предположить, что системы со смешанной инициативой будут работать лучше с точки зрения пользователя. Тем не менее, исследования показали, что пользователи предпочитают системы речевого диалога с активным режимом, поскольку в них выше уровень достижения цели [3]. Другие исследования показали, что инициативные системы имеют большую производительность с неопытными пользователями. Версия той же системы со смешанной инициативой показала худшие результаты для пользователей любого уровня. Эти результаты свидетельствуют о том, что люди адаптируют свое поведение, потому что знают, что они взаимодействуют с машиной, и идут на некоторые ограничения в диалоге для того, чтобы достичь цели (например, заказать билет на самолет или получить справку о ближайших рейсах).

Немаловажную роль в управлении диалогом играет *определение уровня доверия системы.* При этом выполняется контроль за возможными ошибками, сделанными подсистемами, отвечающими за обработку входных данных. Действительно, иногда полезно вводить некоторые подтверждения или проверки

Труды СПИИРАН. 2006. Вып. 3. Т. 1. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2006. Issue 3. V. 1. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

поддиалогов с целью повышения уровня уверенности системы. Заметим, что такие поддиалоги есть и в человеческих диалогах, они позволяют устранить различные виды непонимания (как речевого сигнала, так и смысла). Тем не менее, выбор стратегии подтверждения не тривиален. Разработчик должен решить два основных вопроса: когда и как система должна включать подтверждение поддиалога? Не существует простых ответов на эти вопросы, поскольку различные исследователи приходят к различным выводам.

Прежде всего, разработчик должен решить, когда включать подтверждение поддиалога. Некоторые разработчики утверждают, что это нужно пользователю для того, чтобы обнаружить проблему понимания и что найденная информация всегда должна получить подтверждение [6]. С другой стороны, объективное измерение уровня доверия системы может помочь принять решение. Некоторые системы используют уровень доверия подсистемы автоматического распознавания речи [7], когда для принятия решения необходимо получить от пользователя подтверждение правильности распознавания входной фразы.

Затем идет вопрос, как на каждом уровне можно получить информацию о неуверенности системы. Более того, поскольку полная система диалога имеет подсистему понимания естественного языка, то необходимо учитывать также доверие смыслу и контексту, для того чтобы улучшить общее доверие и решить: какой вопрос следует задать для подтверждения? [8]. Если выявлена необходимость в подтверждении, то МД может выбрать между двумя возможными методами подтверждения:

- явное подтверждение: когда система просит пользователя подтвердить верно распознанную информацию;
- неявное подтверждение: когда информация, нуждающаяся в подтверждении, комбинируется с запросом, связанным с получением следующей части информации. Эта стратегия основана на инстинктивной реакции пользователя на любую некорректную информацию.

Явное подтверждение, как оказалось, работает надежнее. Однако необходимо учитывать, что эта надежность обеспечивается за счет естественности и эффективности взаимодействия, поскольку увеличивает длину диалога. С другой стороны, неявное подтверждение может вызвать путаницу, что приведет к потере надежности управления пользователем. А при условии, что информация распознана корректно, неявное подтверждение имеет существенные преимущества по скорости и удобству взаимодействия.

В общем, выбор между одной из двух стратегий все еще зависит от уровня доверия входной информации. Большинство систем использует неявное подтверждение, когда система имеет достаточно хороший уровень доверия информации. Явное подтверждение используется при низком уровне доверия [9].

Проблема оценки эффективности явных подтверждений является предметом отдельных исследований. Очевидно, что они должны иметь вопросительную интонацию, а не утвердительную и что детали, которые необходимо подтвердить должны быть в конце подтверждающего сообщения: они не должны следовать за вопросом «Это верно?» Поскольку пользователь довольно много успевает сказать до или в течение такого вопроса [6].

В данном разделе были рассмотрены основные стратегии управления диалогом. Далее рассмотрим основные методы оптимизации стратегии диалога.

Труды СПИИРАН. 2006. Вып. 3. Т. 1. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2006. Issue 3. V. 1. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

4. Методы обучения стратегии диалога

Возвращаясь к началу статьи, напомним, что диалог рассматривается как процесс взаимодействия между двумя агентами. Это последовательный процесс, в течение которого два агента пробуют достигнуть цели в рамках определенной задачи: ведут проблемно-ориентированный диалог. Чтобы достичь взаимной цели, агенты обмениваются устными сообщениями, содержащими намерения, до тех пор, пока задача не закончена или один из агентов не отказался продолжать диалог. По аналогии с диалогом между людьми диалоги, ориентированные на конкретную задачу и направляемые целью, должны иметь возможность оценивания взаимодействия и стратегий, которые использует каждый из агентов. Следовательно, должна быть возможность обучения стратегии согласно критерию оптимальности. Все же, не всегда легко определить показатели качества диалога и получить такой критерий. Кроме того, получение речевых данных (реплик реальных диалогов) для обучения стратегии проблемно-ориентированного диалога — это очень сложная задача, требующая оптимизации.

На сегодняшний день разработаны многочисленные методы обучения, благодаря исследованиям, которые проводились в области искусственного интеллекта более половины столетия. Некоторые из методов используют парадигму обучающих алгоритмов — это методы сравнения с эталоном, принятия решения и т.д. Все эти методы объединяет то, что достаточно сложно найти аналитическое решение и поэтому используются статистическое моделирование и мягкие вычисления.

Можно выделить два основных класса методов обучения: управляемое и неуправляемое обучение (или с учителем и без него). В управляемом обучении агенту предъявляются пары: входные данные — решение. При этом обучаемый агент пытается обеспечить правильные решения для новых подобных входных данных. Нейронные сети широко используются как обучающие агенты в методах сравнения эталонов при распознавании речи или рукописного текста. Для обучения модели диалога широко применяются методы стимулирующего обучения. В качестве входных данных используются ситуации, а выходных — ответные действия диалоговой системы. Далее рассмотрим обучение стратегии диалога с учетом методов управления диалогом.

В рамках задачи обучения стратегии, модель диалоговой грамматики не подходит, так как цель алгоритмов стимулирующего обучения состоит в том, чтобы подобрать определенные ситуации к действиям, а не получать общие правила о том, что сделать в общих случаях. Другие способы, такие как сценарная модель или теория диалоговых игр подходят больше, но они не рассматривают диалог как единое целое. Каждый переход в диалоге является частью плана, но при этом не делается никаких попыток исправить ход поддиалога. Таким образом, единственным приемлемым подходом остается метод совместной активности, в котором проектировщик пробует оптимизировать полный диалог, принимая во внимание возможные недоразумения, вероятные ошибки модулей распознавания и понимания речи и т.д. В этом случае диалоговая система может также рассматриваться как агент, пытающийся оптимизировать каждый диалог в целом шаг за шагом согласно поступающим устным сообщениям пользователя.

Для представления моделей диалога наиболее подходят марковские сети, позволяющие описать процесс в виде графа. Применяя методы стимулирующе-

Труды СПИИРАН. 2006. Вып. 3. Т. 1. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2006. Issue 3. V. 1. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

го обучения к проблемам обучения оптимальной стратегии диалога, требуется определить диалог как марковский процесс, который описан в терминах состояний, действий, оценок и стратегии [4]. При этом действия будут определяться экспертом в зависимости от специфики рассматриваемой задачи речевого диалога, а пространство состояний и оценочные параметры будут вырабатываться по взаимодействию с окружающей средой. Действительно, параметры сети, имеющие численные значения, будут постепенно настраиваться по мере поступления данных, а пространство состояний будет построено благодаря упорядочиванию событий, классифицированных модулем распознавания и удовлетворяющих модели задачи.

Разумеется, существует много способов построения пространства состояний [3], и часто утверждается, что пространство состояний сильно зависит от решаемой задачи. Все же, некоторые общие соображения следует принять во внимание:

- каждое состояний должно содержать достаточно информации об истории диалога, чтобы соблюдалось свойство однородности сети;
- пространства состояний часто рассматриваются как информационные в том смысле, что они построены благодаря информации, которую менеджер диалога может извлечь из окружающей среды;
- описание состояния должно содержать достаточно информации для того, чтобы дать точное представление о ситуации, с которой должно быть связано действие;
- пространство состояний должно быть как можно меньше, так как скорость работы алгоритмов стимулирующего обучения пропорционально числу состояний в марковской сети.

Процесс обучения стратегии диалога реализуется следующим образом. На рис. 2 представлена общая структура этого процесса.

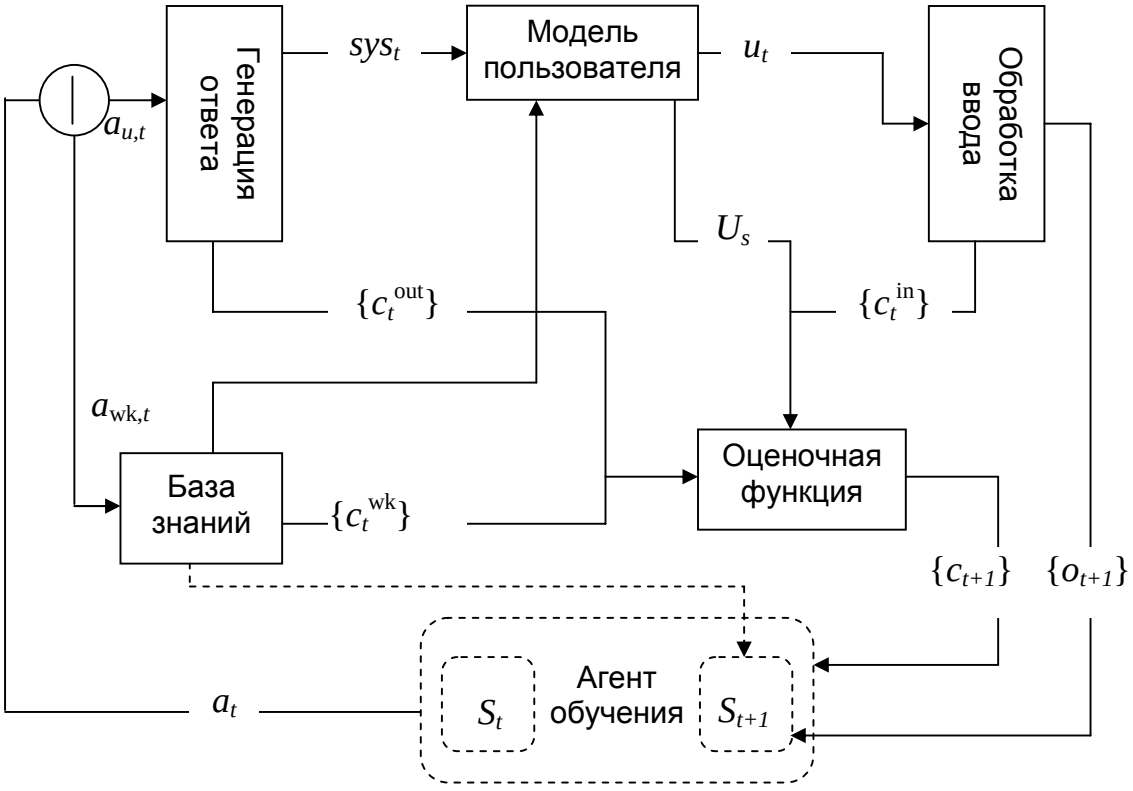

Рис. 2. Процесс обучения стратегии диалога.

Труды СПИИРАН. 2006. Вып. 3. Т. 1. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2006. Issue 3. V. 1. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

Для построения пространства состояний следует учесть следующие данные. Во-первых, известно параметрическое представление задачи. Во-вторых, модуль обработки входных данных обеспечивает данные о наблюдениях. В-третьих, модель пользователя может предоставлять информацию о состоянии пользователя и уровне экспертизы.

Набор действий определяет все возможные ответные реакции, которые сможет выполнить обученная система диалога. Следует отметить, что действия могут быть различных типов. Например, существуют такие вспомогательные элементарные действия, как генерация подсказки, загрузка грамматики речи, начало распознавания речи и т.д. С другой стороны, начало диалога и следование определенной стратегии до конца взаимодействия — это тоже действия. В начале обычно выполняется несколько элементарных действий, не требующих оптимизации очередности исполнения.

Кроме того, выделяются два типа действий в зависимости от объекта их воздействия: действия направлены на пользователя или окружающий мир. Последовательность элементарных действий, направленных на пользователя, обычно начинается с подсказки и строится на базе теории речевых актов. Например, пользователю задается некоторый вопрос, затем эта же информация поступает в блок распознавания речи, чтобы сформировать там необходимую высокоуровневую информацию и уже ожидать один из наиболее вероятных ответов пользователя. Получив от пользователя устное сообщение, система обрабатывает его и определяет переход в новое состояние согласно полученной информации от пользователя. Так как эти последовательности действий зафиксированы в схеме обучения стратегии, то задача эксперта — определить все возможные действия. Кроме того, существуют некоторые общие действия, которые используются большинством диалоговых систем.

Приветствие. Действие, с помощью которого система сообщает пользователю о своих способностях и его возможностях по работе с системой (обычно строится в виде открытого вопроса). Во время этого действия производится общая инициализация диалоговой системы: загрузка словарей, баз данных и знаний, необходимых для распознавания речи, обработки текста и генерации адекватных ответных действий. Например: Система: «Добро пожаловать в нашу систему заказа билетов на поезд! Как я могу Вам помочь?».

Уточняющий вопрос. Действие, с помощью которого система ограничивает предметную область диалога, тем самым, направляя пользователя на конкретную проблему и сокращая набор возможных ответов. В принципе, это действие направлено на повышение точности работы модуля обработки входных сообщений. Чем быстрее система сможет понять намерение пользователя (суть запроса), тем скоре сможет выработать соответствующее ответное действие (выдать интересующую информацию), и в итоге достичь конечной цели диалога. Например, Система: «Пожалуйста, сообщите название пункта прибытия».

Открытый вопрос. Действие, которое может быть исполнено как системой, так и пользователем, направлено на запрос дополнительной информации. При этом не предъявляются строгие ограничения на содержании ответа. Например, Система: «Как я могу помочь Вам?»

Явное подтверждение. Действие, с помощью которого система просит подтвердить определенную информацию, полученную на основе предыдущего сообщения пользователя. Обычно такое действие требует простых ответов

Труды СПИИРАН. 2006. Вып. 3. Т. 1. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2006. Issue 3. V. 1. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

«Да» или «Нет». Например, Система: «Вы сказали, что Вы хотите билет в Санкт-Петербург, правильно?»

Неявное подтверждение. Действие, которое может быть исполнено как системой, так и пользователем, содержит неявный запрос на подтверждение полученной ранее информации, а также запрос новой информации. Это делается путем повторения информации, извлеченной из предыдущего сообщения. При этом ожидается, что агент инстинктивно заметит неправильную информацию и поправит ее или, в случае отсутствия ошибок, сообщит новые данные. Например: Система: «Когда Вы хотите поехать в Санкт-Петербург?»

Окончательное подтверждение. По окончании диалога или поддиалога, система может попросить подтвердить все переменные, полученные в ходе диалога. Например, Система: «Пожалуйста, подтвердите свой запрос: Вы запрашивали билет в одну сторону из Москвы в Санкт-Петербург на следующую субботу».

Изменение условий. В случае невозможности выполнения действия по запрашиваемых условиям, производится отказ и предложение изменить некоторые параметры запроса. Это имеет место, когда система, обработав входные данные и сформулировав запрос к своим информационным базам данных, не смогла найти удовлетворительного ответа. Например, Система: «На поезда из Москвы в Санкт-Петербург на следующую субботу мест нет. Хотели бы Вы изменить дату отправления?»

Восстановление поддиалога. Иногда, диалог проходит затруднительно, так как ответы пользователя часто бывают не последовательны, а также вследствие ошибок распознавания/понимания речи. В этом случае возникают ситуации, когда система принимает решение о необходимости начать диалог заново или с какого-то определенного поддиалога. Такой поддиалог может содержать простую подсказку пользователю для ответа на последний вопрос или может быть более сложным, чтобы оценить насколько пользователь и система понимают друг друга. Например, Система: «дата отправления может, начинаться с завтрашнего дня и до 1 января 2007 года».

Утвердительная подсказка. Система передает информацию пользователю, чтобы обновить его знания. При этом система не ожидает в ответ никакой специальной реакции пользователя и продолжает оставаться в текущем состоянии поддиалога (чаще всего в начале диалога). Это действие направлено на быстрое оказание помощи. Например, Система: «Поезд, который Вы запросили, прибывает на 2 путь».

Окончание диалога. В конце сессии диалога система сообщает об завершении диалога и переходит в изначальное состояние для работы с новым пользователем. Например, Система: «Ваш билет будет доставлен Вам сегодня вечером. Спасибо, что Вы воспользовались автоматизированной системой заказа билетов».

Действия, адресованные системе, обычно ограничиваются типичными вопросами, которые заранее содержатся в системе, и на которые существуют заранее подготовленные формы ответов. В зависимости от запросов пользователя меняется только содержание ответного действия. При этом общая структура действий остается неизменной.

В заключение этого раздела отметим, что разработка диалоговых систем является комплексной задачей и при разработке структуры диалога, прежде всего, учитываются знания экспертов предметной области. При этом не только структура диалога, но и сами фразы, задаваемые диалоговой системой, долж-

Труды СПИИРАН. 2006. Вып. 3. Т. 1. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2006. Issue 3. V. 1. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

ны быть тщательно подобраны, чтобы сократить время диалога до минимума. Так как цель многих диалоговых систем обеспечить пользователя интересующей информацией, то возможно использовать в качестве критерия оптимальности стратегии диалога, среднее время успешного взаимодействия, т.е. когда пользователь проходит все этапы диалога, и по завершению получает необходимую информацию.

В следующем разделе, чтобы получить общее представление о состоянии дел в области диалоговых систем кратко рассмотрим некоторые современные системы речевого диалога.

5. Современные диалоговые системы

Одним из самых перспективных направлений применения диалоговых систем является разработка новых сервисов, систем и услуг, которые могли бы максимально использовать коммуникационные способности человека. Использование речевых технологий особенно актуально в телекоммуникационных приложениях. Впервые технологии распознавания речи были использованы в телекоммуникационных сетях около 10 лет назад. Необходимость внедрения технологий распознавания речи в телекоммуникациях была обусловлена двумя причинами [10]: уменьшение стоимости услуг за счет автоматизации функций сопровождающего персонала и получение дополнительных доходов за счет предоставления пользователю дополнительных интеллектуальных сервисов, которые не использовались ранее из-за высокой стоимости.

Применение речевых технологий в сфере телекоммуникаций стало следующей ступенью развития телекоммуникационных услуг. Среди многих примеров телекоммуникационных сервисов, которые были использованы для автоматизации процесса обслуживания клиентов можно выделить следующие [11]:

- автоматизация операторских функций. Например, такие системы как Voice Recognition Call Processing (VRCP) фирмы AT&T или Automated Alternate Billing System (AABS) фирмы Bell Northern Research позволили автоматизировать часть операторских функций. Система VRCP способна обрабатывать такие вызовы как: составление квитанций и счетов, запрос визитных карточек. VRCP в настоящее время используется в США, обрабатывая свыше миллиарда запросов в год;
- автоматизация справочной системы. Были разработаны системы помощи операторам справочных систем для задачи определения номера телефона в ответ на речевой запрос пользователя. Компании NYNEX и Nortel разработали свои собственные системы, которые могли распознавать названия городов для облегчения процесса поиска населенного пункта на карте. Системы, разработанные Bellcore и Ameritech (система ACNA) позволили клиенту телефонной компании получать название некоторой организации (например, магазина) и ее адрес, продиктовав системе номер телефона интересующей организации;
- информационные сервисы в автоматизированных call-центрах. Существующие Call-центры позволяют абонентам получать при помощи телефона доступ к информации о результатах спортивных соревнований, прогнозе погоды, осуществлять заказ билетов, бронирование номера в гостинице и т.д. В настоящее время начинают внедряться автоматизированные Call-центры, которые могут поддерживать некоторый интеллектуальный рече-

Труды СПИИРАН. 2006. Вып. 3. Т. 1. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2006. Issue 3. V. 1. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

вой диалог с абонентом. Так, например, система American Airlines Dial-a-Flight позволяет автоматизировать продажу и заказ авиа билетов, а также получение справочной информации о рейсах по телефону. По данным на 2002 год около 81% американских компаний имеет собственный или арендует внешний call-центр, в которых занято свыше 1,55 млн. операторов [12];

- голосовые банковские услуги. В начале 90-х годов в Японии фирма NTT разработала систему ANSER (Automatic Answer Network System for Electrical Requests), которая обеспечивает речевой интерфейс для доступа к кредитным картам, банковским счетам клиентов, сообщает остатки на счете, позволяет осуществлять перевод средств с одного счета на другой, и т.д. В настоящее время система ANSER установлена более чем в 70-и японских городах и обслуживает более 600 банков. Система ежегодно обрабатывает порядка 360 миллионов звонков;
- голосовые порталы. Голосовой портал — это новый и специфический тип Интернет-портала. Он предоставляет пользователям возможность получать и управлять информацией, размещенной в Интернет, при помощи голоса. Абонент может использовать необходимый ему сервис как посредством телефона, так и голосового Web-броузера, которые, используя протоколы WAP или http, могут получить доступ к голосовому порталу и соответственно базам данных с различной информацией. На основе существующих интеллектуальных технологий фирмами TellMe, BeVocal, Webversant и др. разработаны голосовые порталы для различных прикладных задач. По данным аналитической группы Allied Business Intelligence в США в 2001 существовало около 4 миллионов голосовых порталов, а к 2005 году их число составило почти 17 миллионов.

Кроме телекоммуникационных приложений существует масса задач, где речевые технологии могли бы быть полезны. Для того чтобы классифицировать круг задач, где сегодня применяются диалоговые системы, рассмотрим табл. 1.

Таблица 1

Основные типы прикладных систем речевого диалога

Цель приложения	**Выполняемые функции или предоставляемые услуги**
Альтернатива клавиатурному вводу	Стенографирование, генерация документов.
Доступ к информации и передача речевых сообщений по телефону	Доступ к Интернет ресурсам. Осуществление банковских переводов. Использование естественного речевого диалога. Голосовая почта. Мгновенное обновление трафика, для выбора маршрута. Бесконтактное управление мобильным телефоном.
Образование	Обучения детей произношению звуков речи и различных слов. Обучения иностранным языкам взрослых — машины часто произносят слова более точно, чем учитель — не носитель языка.
Исправление дефектов речи	Демонстрация движения языка при произнесении речи, сгенерированная на компьютере.
Решение слуховых и визуальных проблем	Чтение текстов для людей со слабым зрением. Синтез речи по жестам. Обучающие системы интерактивного аудио-визуального взаимодействия для детей с отставаниями в развитии.
Безопасность	Верификация и идентификация диктора по голосу.
Развлечение	Игрушки разговаривают уже много лет, но сейчас они уже понимают, кто с ними говорит и могут приветствовать Вас по имени.

Труды СПИИРАН. 2006. Вып. 3. Т. 1. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2006. Issue 3. V. 1. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

Приведем несколько примеров наиболее успешных систем речевого диалога. Технологии речевого взаимодействия сегодня наиболее активно развиваются в области биологии и медицины, поскольку именно эти области обладают детально специфицированными моделями знаний.

Сервис по уходу за диабетиками Dias-Net-PN (Diabetes Advisory System — Personal Networked) консультационная система, которая предоставляет помощь в диагностике диабета, а также лечении болезни и ее осложнений [13]. На рис. 3 представлена общая структура пользователей системы. Система обеспечивает взаимодействие медицинских работников и пациентов посредством компьютера или мобильного телефона. Проведенные исследования показали высокую эффективность сети Dias-Net-PN. Были найдены решения ориентированные на специальные нужды пользователей.

Рис. 3. Консультационная система Dias-Net-PN.

Другая *медицинская система Medical Studio* основана на многомодальной платформе. Эта платформа поддерживает полную поточную обработку, в том числе обработку входных данных, визуальное наблюдение, управление предоперационным планированием и управление во время операции. На основе данной платформы могут создаваться системы помощи проведения различных операций на основе аудиоинформации, видео-наблюдения и автоматической обработки данных. Компоненты данной платформы, разработанные в сотрудничестве нескольких исследовательских центров и медицинских клиник, за счет своей универсальности могут широко применяться в различных областях медицины [14].

Автоматизированный многоязычный call-центр AMITIES (Automated multilingual interaction with information and services) разработан при поддержке 5 европейской рамочной программы и американского агентства DAPRA [15]. Проект

Труды СПИИРАН. 2006. Вып. 3. Т. 1. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2006. Issue 3. V. 1. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

создавался как крупномасштабная, экспериментальная диалоговая (в человекоподобной форме) система с использованием информационного материала, полученного по фактическим и целенаправленным диалогам типа человек–человек. Оказалось, что получение реальных данных в достаточных количествах — это длинный и сложный процесс, который затрудняется, в том числе, юридическими препятствиями.

Система бронирования авиабилетов Mercury разработана в 2000 году [16]. Она обеспечивает доступ к базе данных полетов в режиме реального времени и позволяет пользователю планировать поездку, а также узнавать стоимость билетов в крупнейших аэропортах мира. Управление стратегией диалога основано на наборе упорядоченных правил механизма диалогового взаимодействия. В данный момент система активно разрабатывается, а отдельные модули, разработанные в рамках проекта Mercury, такие, как интерпретация даты и времени, могут использоваться в других приложениях.

В недалеком будущем пользователям будут также доступны такие новые интеллектуальные сервисы как [11]:

- технология интеллектуальных агентов. AT&T ведет разработку системы Wildfire, которая по замыслу разработчиков должна предоставить возможность клиентам вести речевой диалог с интеллектуальным агентом для управления телефонными звонками, сообщениями, специальными клиентскими сервисами, а также иметь способность адаптироваться к предпочтениям и требованиям пользователя;
- системы предоставления помощи клиенту. Целью таких систем является организация диалога с пользователем для предоставления ему необходимой помощи по работе некоторой системы, которые позволяют избежать работы со сложными конструкциями всевозможных меню или избавляют от необходимости освоения сложной терминологии конкретной системы;
- системы диктовки речи для создания и обработки различных документов;
- многомодальный интеллектуальный помощник. Фирмы Sun Microsystems, Nuance и BeVocal поставили задачу разработать новую многомодальную (объединяющую речь, тактильную, визуальную информацию, а также, возможно, и жесты) архитектуру/систему VoiceTone для мобильных и настольных компьютеров, а также телефонов, которая сможет стать настоящим помощником человека, общаясь с ним в естественной для человека форме.

В основе всех вышеперечисленных существующих систем человеко-машинного диалога заложены основные принципы, изложенные в разделах 2, 3 и 4. В данном разделе были рассмотрены существующие диалоговые системы, а также перспективы применения речевых технологий. Тенденции, появившиеся в последние годы, свидетельствуют о том, что существующие технологии распознавания речи позволяют создавать полезные и эффективные приложения. Однако, существует ряд проблем, которые еще пока не позволяют охватить желаемые области применения речевых технологий.

6. Заключение

На сегодняшний день в области речевых технологий на первый план выходят приложения для новых прикладных областей, таких как сотовая связь, IP-телефония, Интернет и других, а также специфические приложения, предназначенные для людей с ограниченными возможностями, инвалидов и больных.

Труды СПИИРАН. 2006. Вып. 3. Т. 1. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2006. Issue 3. V. 1. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

Существующие модели понимания речи пока еще значительно уступают речевым способностям человека, что свидетельствует об их недостаточной адекватности и ограничивает применение речевых технологий в промышленности и быту. Для решения глобальной проблемы человеко-машинного взаимодействия стали использовать дополнительные виды каналов передачи информации (речь, артикуляция губ, жесты, направление взгляда и т.д.). В результате появились, так называемые, многомодальные интерфейсы [1].

Такие интерфейсы позволяют обеспечить наиболее эффективное и естественное для человека взаимодействие с различными автоматизированными средствами управления и коммуникации. Однако, появление новейших технологий, использующих распознавание речи, еще не является показателем развития этого научного направления. Скорее развитие объясняется нахождением новых способов использования старых методов. Не следует забывать, что и без помощи дополнительных модальностей человек способен вести полноценный речевой диалог, располагая инструментами устранения неоднозначности и непонимания.

Основной проблемой проектирования систем речевого диалога является отсутствие формальных методов параллельной обработки различных элементов речи. Хотя и были различные попытки применить методы и языки программирования для параллельной обработки в языковых технологиях, результаты все еще не убедительны. Для прогресса в языковых технологиях крайне необходимо более детально проработать различные формальные подходы на основе строгих математических моделей.

На каждом уровне обработки потока информации существуют свои актуальные проблемы. На уровне семантического анализа это проблема разумной степени полноты входной информации. Существует некоторый разумный уровень полноты диалоговых фраз, в принципе достаточный для их надежного понимания без переспросов. Поэтому статистические данные, с достаточной полнотой отражающие речевое поведение человека, могут быть полезными в решении данной проблемы.

На уровне управления диалогом стоит выделить проблему построения оптимального менеджера диалога, которая требует решения целого комплекса задач по моделированию диалога, разработке стратегии диалога, определению уровня активности пользователя и системы, определению уровня уверенности системы, а также разработке методов обучения стратегии диалога. Решение всех этих задач требует проработки методов искусственного интеллекта, а также кропотливой и объемной работы с экспертами и опытными статистическими данными.

В данной работе очерчен круг проблем, связанных с проектированием диалоговых систем, в частности, рассмотрены основные задачи построения оптимального менеджера диалога, а также приведены примеры реальных систем речевого диалога, основанных на различных подходах.

Литература

1. *Ронжин А Л, Карпов А. А*. Многомодальные интерфейсы: основные принципы и когнитивные аспекты // Труды СПИИРАН. Вып. 3, т. 1. СПб.: Наука, 2006. [В этом же томе].
2. *Bengios S*. Multimodal Speech Processing Using Asynchronous Hidden Markov Models // Information Fusion. 2004. P.81–89.
3. *Pietquin O*. A Framework for Unsupervised Learning of Dialogue Strategies. UCL presses, 2004. 246 p.

4. *Levin E., Pieraccini R.* A Stochastic Model of Computer-Human Interaction for Learning Dialogue Strategies // Proc. of the 5 European Conference on Speech Technologies (Eurospeech'97). Rhodes, Greece, 1997. P. 1883–1886.
5. *Smith R , Hipp R, and Alan W. Biermann.* A Dialog Control Algorithm and Its Performance / Bates, Madeleine and Oliviero Stock (eds.) // Third Conference on Applied Natural Language Processing, 31 March -3 April 1992. P. 9–16.
6. *Mclnnes F., Nairn I., Attwater D., Edgington M., Jack M.* A Comparison of Confirmation Strategies for Fluent Telephone Dialogues // Proceedings of the 17th International Symposium on Human Factors in Telecommunication (HFT'99). 1999. P. 81–89.
7. *Williams G., Renals S.* Confidence Measures for Hybrid HMM/ANN Speech Recognition // Proceedings of the 5th European Conference on Speech Technology, (Eurospeech'97) Rhodes, 1997. P. 1955–1958.
8. *Komatani K., Kawahara T.* Generating Effective Confirmation and Guidance Using Two-level Confidence Measures for Dialogue Systems // Proceedings of the 8th International Conference on Spoken Language Processing (ICSLP'00). 2000. Vol.2. P. 648–651.
9. *Bouwman G., Sturm J., Boves L*. Incorporating Confidence Measures in the Dutch Train Timetable Information System Developed in the ARISE project // Proceedings of the International Conference on Avoustics, Speech and Signal Processing (ICASSP), 1999. Vol.1. P. 493–496.
10. *Rabiner L. R.* Applications of Speech Recognition in the Area of Telecommunications // 1997 IEEE Workshop on Automatic Speech Recognition and Understanding Proceedings, 1997. P. 501–510.
11. *Иванова Т. И.* Компьютерные технологии в телефонии. М: Эко-Трендз, 2002. 300 с.
12. *Крестьянинов С. В*. Интеллектуальные сети и компьютерная телефония. М.: Радио и связь, 2001. 238 с.
13. *Pedersen C. F., et. al.* Analysis and Design of a PN based Health Care Service for Diabetics // Proceedings of workshop on "My Personal Adaptive Global NET: Visions and beyond", Shanghai China November, 2004. Compact disc proceedings.
14. *Gemo M., Kitney R.* Medical applications. Similar Dreams. Multimodal Interfaces in Our Future Life. presses universitaires de Louvan, 2005. P.63–75.
15. *Hardy H., Strzalkowski T. and Wu M.* Dialogue Management for an Automated Multilingual Call Center // Proceedings of HLT-NAACL 2003 Workshop: Research Directions in Dialogue Processing, Edmonton, Canada, June 2003. P. 10–12.
16. *Seneff S. and Polifroni J.* Dialogue Management in the MERCURY Flight Reservation System // Proc. ANLP-NAACL 2000 Satellite Workshop, Seattle, May 2000. P. 1–6.