

Инь Шуай, аспирант, shuai.yin@yandex.ru, А. С. Ющенко, д-р техн. наук, проф., yusch@bmstu.ru, Московский государственный технический университет им. Н. Э. Баумана, г. Москва

Диалоговая система управления роботом на базе теории конечных автоматов¹

Рассматривается система речевого диалогового управления манипуляционными роботами. Проведен анализ основных методов автоматического распознавания речи, понимания речи, управления диалогом, синтеза голосовых ответов в диалоговых системах. Рассмотрены три типа управления диалогом: "инициатива системы", "инициатива пользователя" и "комбинированная инициатива". Предложена система объектно-ориентированного диалогового управления роботом на основе теории конечных автоматов состояния с использованием глубокой нейронной сети. Основное отличие предложенной системы заключается в предварительном выполнении диалога, в процессе которого робот может выполнять движения, направленные на получение дополнительной информации. После этого робот выполняет поставленную задачу в автоматическом режиме. Такой способ построения диалогового управления роботом позволяет автоматически корректировать результат распознавания речи и соответствующих действий робота и выполнять диалоговое управление в темпе, близком к темпу работы хирурга с человеком-оператором.

Управление роботом возможно в двух режимах. Специальный режим дает возможность непосредственно управлять движениями манипулятора голосовыми командами пользователей. Общий режим расширяет возможности оператора, позволяя ему получить дополнительную информацию в реальном времени.

Необходимость коррекции результата распознавания речи и выполнения действий робота может быть вызвана особенностями речи пользователя, помехами в информационной системе или некорректными голосовыми командами.

Процесс коррекции состоит из трех этапов. На первом этапе выполняется непрерывное преобразование речи в текст в реальном масштабе времени с использованием глубокой нейронной сети, учитывающей особенности и скорость речи различных пользователей. На втором этапе осуществляется управление диалогом на основе теории конечных автоматов. Наконец, на третьем этапе осуществляется управление действиями робота с учетом его текущего состояния.

В целях реализации диалога между пользователем и роботом, близкого к естественному по темпу и по содержанию, создается база сценариев возможных диалогов.

В проведенных экспериментах разработанная диалоговая система использовалась для управления манипулятором KUKA. Диалоговая система реализована в среде Python. Управление роботом осуществлялось с помощью программного обеспечения RoboDK. Результаты экспериментов подтвердили работоспособность и эффективность диалоговой системы управления роботом. Получена достаточно высокая точность распознавания (92 %); при этом скорость автоматического распознавания речи позволяла вести диалог в темпе, близком к темпу естественной речи.

Ключевые слова: глубокая нейронная сеть, скрытая марковская модель, обучение, управление диалогом, конечный автомат состояния, система управления роботом

Введение

Диалоговая система — это средство обмена информацией между человеком и машиной с использованием естественного языка. Такая

система позволяет управлять роботами операторам, не имеющим специальной подготовки. Диалоговая система включает несколько подсистем, в том числе блок автоматического распознавания речи, блок понимания речи, блок управления диалогом, блоки генерации ответов и синтеза голоса [1] (рис. 1).

¹Работа подготовлена при финансовой поддержке гранта РФФИ № 18-07-01313.

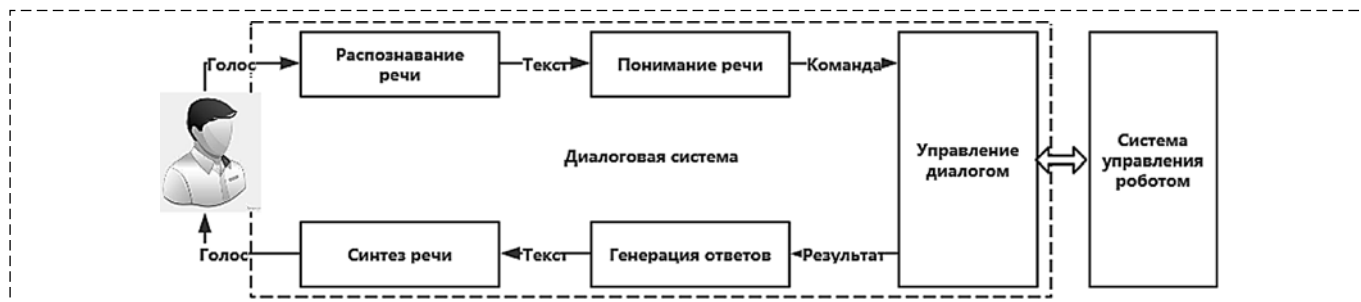


Рис. 1. Схема диалоговой системы управления роботом

Fig. 1. Diagram of the robot interactive control system

Блок распознавания речи преобразует голосовое сообщение в текст. Задача блока понимания речи состоит в том, чтобы правильно интерпретировать это сообщение независимо от формы его выражения. Блок управления диалогом получает информацию от системы управления роботом и управляет диалоговым процессом в соответствии со сценарием диалога. Текст синтезируемых ответов преобразуется в голосовой ответ в блоке "синтез речи".

Диалоговые системы "человек—машина" условно можно разделить на два типа — без ограничений и с ограничениями [2]. Диалоговая система без ограничений, обладая большими базами знаний в различных областях, теоретически позволяет машине ответить на любой вопрос человека. Ограниченная диалоговая система является объектно-ориентированной и разрабатывается для специальной области применения. Несмотря на большое число публикаций и способов организации систем диалогового управления, в каждом конкретном случае возникает проблема наиболее эффективного способа обработки речевых сигналов и управления диалогом с учетом особенностей задачи управления. В статье предлагается структура объектно-ориентированной диалоговой системы для управления манипуляционными роботами.

Методы обработки голосовых сигналов

Предварительным этапом распознавания речи является обработка голосовых сигналов. Обычно голосовые сигналы преобразуются в частотную область с использованием методики MFCC (Mel-Frequency Cepstral Coefficients) [3]. Процедуры распознавания речи можно разделить на процедуры распознавания отдельных слов и распознавание непрерывной речи. Для изолированного распознавания слов применяется акустическая модель динамического искажения времени (DTW — Dynamic Time Warping) [3]. Она чаще и эффективнее применяется без модели конкретного языка. Для непрерывного распознавания речи применяется скрытая марковская модель (НММ — Hidden Markov Model), которая обычно рассматривается совместно с моделью языка.

Напомним, что скрытая марковская модель отличается от марковской модели тем, что скрытая марковская модель определяется тремя параметрами: вероятностью перехода

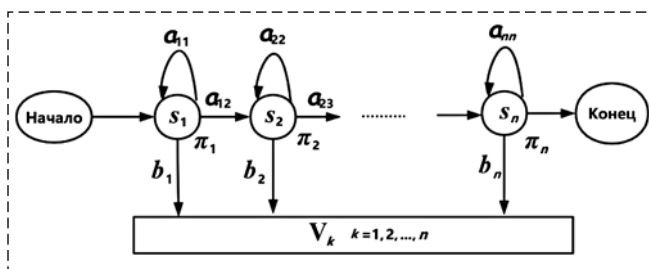


Рис. 2. Схема скрытой марковской модели

Fig. 2. Scheme of the hidden Markov model

между состояниями, вероятностью начального состояния и наблюдаемой вероятностью, в то время как марковская модель построена только по двум первым параметрам. Дискретная НММ представляет собой упорядоченную последовательность состояний (рис. 2) и может быть выражена в форме $\{A, \pi, B\}$, где $A = [a_{ij}]$, $i, j = 1, 2, 3, \dots, N$ — матрица вероятности перехода между N состояниями, $a_{ij} = P(s_t = j | s_{t-1} = i)$, $i, j = 1, 2, 3, \dots, n$; $\pi = [\pi_i]$, $i = 1, 2, 3, \dots, N$ — матрица вероятности начального состояния, $\pi_i = P(s_1 = i)$, $i = 1, 2, 3, \dots, n$; $B = [b_i]$, $i = 1, 2, 3, \dots, N$ — наблюдаемые вероятности $b_i(o_t) = P(o_t = v_t | s_t = i)$, $i = 1, 2, 3, \dots, N$; s_t — состояние в момент времени t ; v_t, o_t — наблюдаемые векторы.

Перечислим три задачи, которые надо решить при использовании скрытой марковской модели:

1. Вычисление наблюдаемых вероятностей $P(o_t | s_t^{(i)})$ при известных параметрах $\{A, \pi, B\}$ и наблюдаемых векторах v_k с помощью алгоритма FBA (Forward-backward algorithm) [4].

2. Обучение скрытой марковской модели с помощью алгоритма Баума-Уэлча (Baum-Welch) [5], в результате чего определяются параметры $\{A, \pi, B\}$ при известных наблюдаемых векторах v_k .

3. Вычисление наиболее вероятного состояния при известных параметрах $\{A, \pi, B\}$ и наблюдаемых векторах v_k с применением алгоритма Витерби (Viterbi) [6].

При распознавании речи состояниями модели являются фонемы, или отдельные слова. Параметры наблюдаемых вероятностей B вычисляются, например, с помощью метода GMM (Gaussian Mixture Models), метода DTW (Dynamic Time Warping) или комбинированного метода GMM-HMM [7, 8].

Для решения задач голосового управления в робототехнике целесообразно приме-

нять глубокие нейронные сети (Deep Neural Network, DNN), которые обладают значительно большими возможностями, чем обычные нейронные сети, от которых они отличаются по числу слоев [9]. В глубоких нейронных сетях имеются не менее пяти скрытых слоев. Однако нейронные сети не способны непосредственно моделировать голосовые сигналы. Поэтому, чтобы использовать способность нейронных сетей к классификации, целесообразно рассмотреть комбинированную модель DNN-HMM (рис. 3).

Модель DNN-HMM, показанная на рис. 3, состоит из глубоких нейронных сетей DNN и скрытой марковской модели HMM. Скрытая марковская модель (HMM) использована для описания динамики голосовых сигналов; вероятности наблюдаемых векторов вычисляются нейронными сетями. Вероятность начального состояния предполагается известной. При заданных характеристиках голосовых сигналов на выходах нейронных сетей будут получены априорные вероятности состояний. Обучение нейронной сети проводится с помощью алгоритма обратного распространения ошибки (Backward Error Propagation — BP). На рис. 3 обозначено: a_{12} , a_{23} — вероятности перехода между состояниями; b_1 , b_2 , b_3 , ..., b_n — наблюдаемые вероятности. Приведенная на рис. 3 сеть DNN имеет шесть скрытых слоев по 2048 нейронов, 440 входных и 1024 выходных сигналов сети. Глубокая нейронная сеть вычисляет вероятность $P(s|\mathbf{o})$, наблюдаемая вероятность $P(\mathbf{o}|s)$ определяется по формуле Байеса:

$$P(\mathbf{o}|s) = \frac{P(s|\mathbf{o})P(\mathbf{o})}{P(s)}, \text{ где } P(s) — \text{априорная веро-}$$

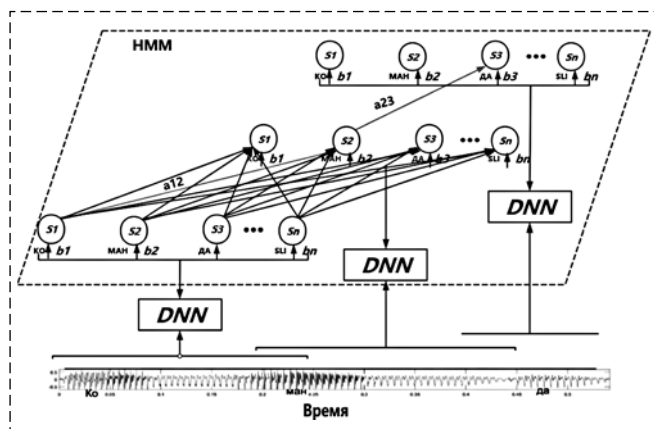


Рис. 3. Схема модели DNN-HMM (распознавание слова "команда")

Fig. 3. DNN-HMM model diagram (recognition of the word "command")

ятность состояния s ; величина $P(\mathbf{o})$ не зависит от последовательности слов и может быть принята равной 1.

При непрерывном распознавании речи рассматриваемая модель DNN-HMM имеет ряд преимуществ по сравнению с моделью GMM-HMM. Она позволяет значительно повысить точность распознавания по сравнению с моделью GMM-HMM. Кроме того, в модели DNN-HMM используется только одна нейронная сеть для вычисления плотности распределения вероятностей состояний $P(s|\mathbf{o})$. В модели GMM-HMM используются модели GMM для вычисления наблюдаемых вероятностей по формуле [10]

$$P(\mathbf{o}_1^N | \mathbf{s}_1^N) = \prod_{n=1}^N b_{s_n}(\mathbf{o}_n) = \prod_{n=1}^N \sum_{m=1}^M \frac{c_{s_n, m}}{(2\pi)^{M/2} [\det(\Sigma_{s_n, m})]^{1/2}} \times \exp \left[-\frac{1}{2} (\mathbf{o}_n - \boldsymbol{\mu}_{s_n, m})^T \Sigma_{s_n, m}^{-1} (\mathbf{o}_n - \boldsymbol{\mu}_{s_n, m}) \right],$$

где $\boldsymbol{\mu}_{i, m}$ — векторы математического ожидания; $\Sigma_{i, m}$ — ковариационные матрицы; $c_{i, m}$ — весовые коэффициенты; $\mathbf{s}_1^T = \{s_1, s_2, \dots, s_T\}$ — последовательность состояний; $\mathbf{o}_1^T = \{\mathbf{o}_1, \mathbf{o}_2, \dots, \mathbf{o}_T\}$ — последовательность наблюдений.

В отличие от модели DTW модель DNN-HMM способна распознавать естественную речь, что особенно важно при разработке робототехнических систем коллаборативного типа, ориентированных на работу с пользователем, не имеющим специальной подготовки.

Поскольку для обучения нейронных сетей нужно иметь большие массивы известных входных и выходных сигналов, то вначале необходимо обучить модель GMM-HMM с использованием алгоритма Viterbi и задать входные и выходные сигналы нейронных сетей с помощью модели GMM-HMM. Таким образом, эффективность глубоких нейронных сетей DNN зависит от выбранной модели GMM-HMM.

Способы распознавания и понимания речи

В качестве основного средства для распознавания речи роботом целесообразно выбрать языковую модель n -gram [11], которая позволяет вычислять вероятность следующего слова по словам, которые появились раньше. Ины-

ми словами, языковая модель n -gram способна вычислить состояние n по предыдущим $n - 1$ состояниям. В модели n -gram обычно имеются модели 1-gram, 2-gram и 3-gram. Модель 1-gram, которая используется наиболее часто, прогнозирует состояние n только по состоянию $n - 1$, модель 2-gram прогнозирует состояние n только по состояниям $n - 1$ и $n - 2$, и т. д.

Задача понимания естественной речи представляет собой интерпретацию смысла сообщения по тексту. В объектно-ориентированной системе задача интерпретации смысла существенно упрощается за счет синтаксиса — правил формирования высказываний. Например, в робототехнической системе структура сообщения о выполнении манипуляционной операции, определяющая синтаксис высказывания, может быть представлена в виде стандартного фрейма: $\langle(1) \text{ название операции} \rangle$, $\langle(2) \text{ объект операции} \rangle$, $\langle(3) \text{ обстоятельство места} \rangle$, $\langle(4) \text{ обстоятельство образа действия} \rangle$ [12]. Например, $\langle(1) \text{ поставить} \rangle$, $\langle(2) \text{ красный короб} \rangle$, $\langle(3) \text{ на большой стол} \rangle$, $\langle(4) \text{ без удара} \rangle$. Обратим внимание на использование в таком высказывании лингвистических переменных, которыми пользуется человек в обычной жизни.

Еще один метод понимания речи рассматривает проблему понимания речи как проблему распознавания образов. При этом наиболее правдоподобный смысл слова или фразы M определяется в результате сравнения с образцами W соответствующих текстов, которые содержатся в результатах распознавания речи на предыдущих этапах. Таким образом, наиболее правдоподобным считается объект, обладающий наибольшей условной вероятностью, что можно записать как $P(M|W)$. Преимущество этого метода — возможность получения наилучшего варианта смысла пользователем из больших массивов данных, полученных на предыдущих этапах обучения. Существуют также методы, основанные на применении марковских двухслойных сетей [13].

Управление диалогом

Управление диалогом непосредственно влияет на то, как естественный и интеллектуальный диалог системы воспринимается пользователями. Задача управления диалогом — определить следующее действие системы, учитывая предыдущий ход диалога. Диалоговый про-

цесс состоит из обращений между субъектом, делающим запрос, и субъектом, отвечающим на запрос, которые зависят от контекста дискурса. В рассматриваемом случае один из субъектов диалога — человек, а другой — робот. Диалоговый процесс можно рассматривать как обмен информацией, в которой инициатива может принадлежать и пользователю, и технической системе. В соответствии с этим можно выделить три вида диалога: 1) система обладает инициативой управления диалогом; 2) пользователь управляет диалогом, и система реагирует на его запросы; 3) система осуществляет общий контроль над диалоговым процессом, но в то же время пользователь может вмешиваться и изменять направление диалога, например, используя указание "отменить задачу А, выполнять задачу Б...". Это диалог смешанного типа.

Существуют несколько методов управления диалогом. Ряд методов использует рекуррентные нейронные сети [14, 15]. В работе [16] применяется рейтинг сообщений. Система управления диалогом может создавать следующие состояния диалога: по его текущему состоянию или по запросу клиента. Существуют подходы к управлению диалогом на основе анализа структуры текста [17] и подходы, предусматривающие обучение диалоговой системы по наблюдаемым записям диалога [18]. Чтобы решить проблемы контроля за состоянием диалога и управления диалогом, можно применить методику конечных автоматов состояния [19]. Такая методика применялась в ряде работ, так, в работе [20] для управления диалогом применялись модифицированные сети Петри. Напомним, что конечным автоматом состояния (КАС) называется математическая модель, которая содержит текущее состояние, условие действия и условие перехода в следующее состояние. Действие будет инициировано или будет выполнен переход одного состояния в другое после выполнения условия перехода. После выполнения действия КАС может перейти в новое состояние или оставаться в исходном состоянии.

На рис. 4 в качестве примера показан конечный автомат, который описывает диалоговую систему управления манипулятором. Здесь приняты следующие обозначения для действий и условий: 1Д (Действие) — инициализация модулей; 2Д — запуск системы; 3Д — инструкции по использованию; 4Д — распознавание речи; 5Д — запуск подсистемы специального режи-

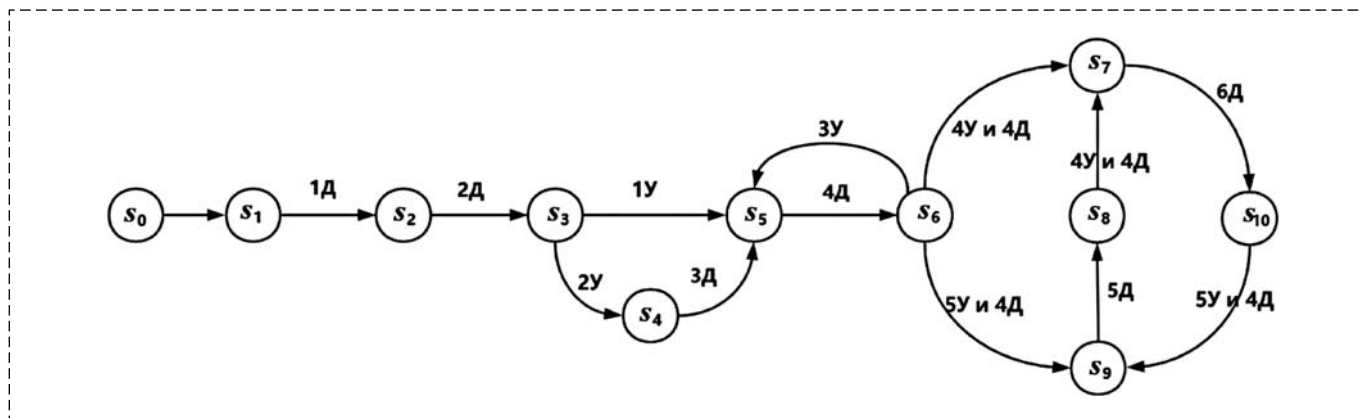


Рис. 4. Конечный автомат состояния диалоговой системы управления манипулятором
Fig. 4. The finite-state automate in interactive manipulator control system

ма; 6Д — запуск подсистемы общего режима; 1У (Условие) — запуск системы первый раз; 2У — запуск системы не первый раз; 3У — команда неправильная; 4У — общий режим; 5У — специальный режим. Это система комбинированного типа. В представленной схеме конечный автомат состояния имеет девять состояний. Первые пять состояний автомат проходит последовательно, переходы по остальным состояниям зависят от запросов пользователей. Система распознает запросы пользователей с помощью алгоритма (DNN-HMM). Пользователь должен ответить на вопрос системы: "вводить специальный режим или общий режим?", "управлять ли положением инструмента?", "введите голосовую команду" и т. д.

На практике в объектно-ориентированной диалоговой системе число входных запросов клиентов ограничено (обычно число запросов от 20 до 30). В ориентированной на "естественную" речь (неограниченной) диалоговой системе предполагается распознавание любых высказываний. Чтобы реализовать такой диалог используются большие массивы данных для обучения. По сравнению с неограниченной диалоговой системой объектно-ориентированная система обладает не только ограничениями, но и рядом преимуществ, в том числе высокой скоростью распознавания речи, высокой эффективностью и простотой обучения.

Система диалогового управления манипуляционным роботом

Ниже рассматривается возможность применения объектно-ориентированной диа-

логовой системы для управления манипуляционным роботом в специальном режиме и в общем режиме.

Специальный режим дает возможность непосредственно управлять манипулятором по голосовым командам клиентов. Например, "положить красный блок на место А", "убрать синий блок" и т. д. В этом режиме применены акустическая модель (DNN-HMM), языковая модель (n -gram) и словарь для распознавания нескольких заранее определенных объектно-ориентированных речевых запросов (рис. 5).

С помощью открытых исходных пакетов Kaldi ASR обучаются акустическая модель DNN-HMM и языковая модель n -gram. Акустическая модель представляет слова русского языка с помощью латинского алфавита. Например, слово "положить" представляется как p o l o z h y u t t, а слово "красный" как k r a a s n y j. В акустической модели на базе глубоких нейронных сетей определяются вероятности возможных состояний (слов). Состояние максимальной вероятности считается достоверным отображением голосовых сигналов. Применение языковых моделей позволяет повысить точность распознавания речи. Например, если вероятность высказывания "положить на место" составляет 0,1, а вероятность высказывания "положить блок" составляет 0,5, то запрос "положить блок" считается достоверным. Анализатор высказываний способен преобразовать их в команды, которые должны быть правильно интерпретированы роботом. Например, высказывание "положить красный блок на место А" преобразуется в последовательность команд: "манипулятор из положения В (полученного при выполнении

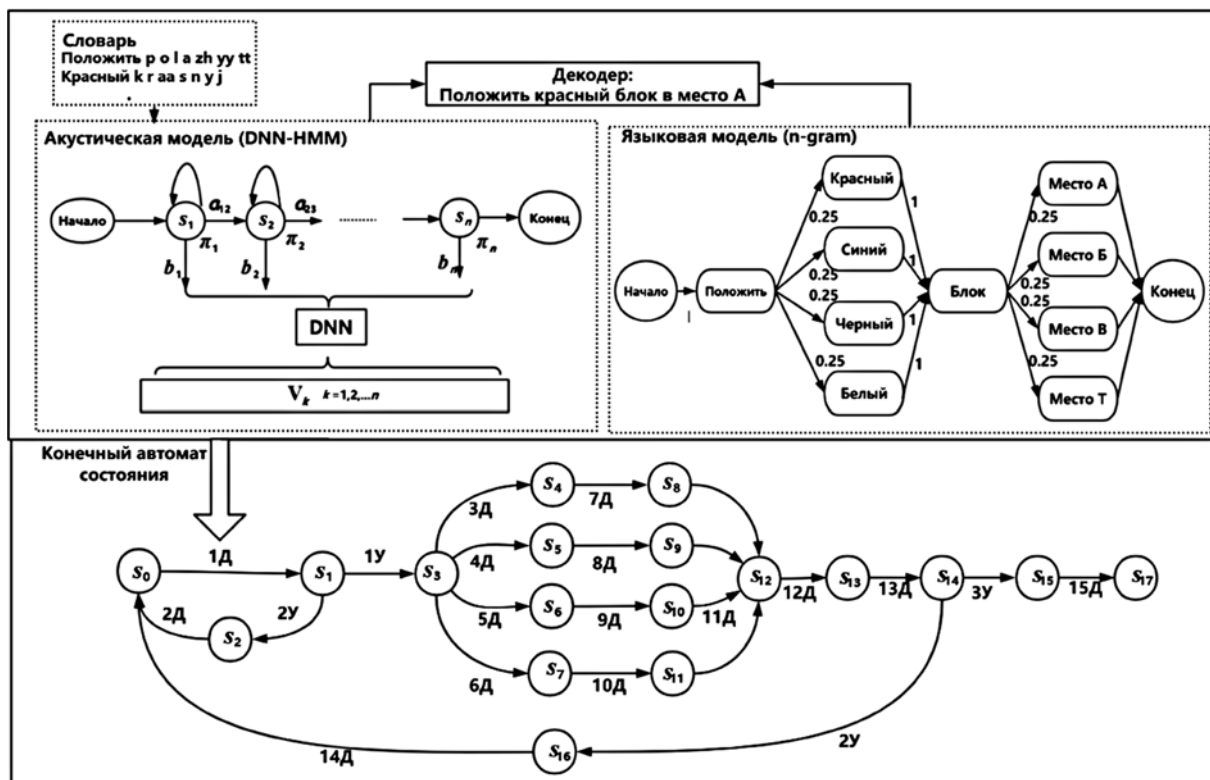


Рис. 5. Схема специального режима. Левая часть графика — метод распознавания русской речи; правая часть графика — управление диалогом на основе теории конечных автоматов состояния

Fig. 5. Scheme of the special mode. The left part of the graphics is the Russian speech recognition method; the right part of the graphic is dialog management based on a finite — state automata

предыдущей команды "взять красный блок") перевести в положение А. Заметим, что в более сложных случаях необходимо решать задачу автоматического планирования действий. Например, если в приведенной выше команде на красном блоке находится предмет С, то его надо сначала убрать [21]. С помощью технического зрения робота можно определить координаты красного блока в положении В и координаты заданного оператором положения А, после чего передать эти координаты в систему управления манипулятором. Для распознавания сложных объектов можно использовать сверточную (конволюционную) нейронную сеть [22]. После выполнения операции предполагается проверка реального состояния, т. е. того, находится ли блок в положении А. Если ответ положительный, то задача считается выполненной.

На рис. 5 приняты следующие обозначения для узлов и ребер графа: 1Д (Действие) — распознавание речи; 2Д — запуск голосового файла: "команда неправильная, снова введите команду"; 3Д — проверка положения "место А";

4Д — проверка направления "в"; 5Д — проверка объекта "красный блок"; 6Д — проверка глагола "положить"; 7Д — обнаружение места А; 8Д — определение направления движения схвата манипулятора; 9Д — обнаружение красного блока; 10Д — формирование траектории схвата манипулятора; 11Д — определение координат захвата; 12Д — схват манипулятора движется по траектории; 13Д — обнаружение красного блока; 14Д — запуск голосового файла "операция не выполнена, снова ввести команду"; 15Д — запуск голосового файла "операция выполнена, ввести новую команду"; 1У (Условие) — команда правильная; 2У — команда неправильная; 3У — красный блок не находится в нужном месте.

Заметим, что в специальном режиме не выполняется синтез речи, все голосовые ответы содержатся в заранее подготовленном файле записи, например, "операция выполнена", "введите голосовую команду" и т. д.

Общий режим расширяет возможности оператора, позволяя ему получить дополнительную информацию в реальном времени. Наибо-

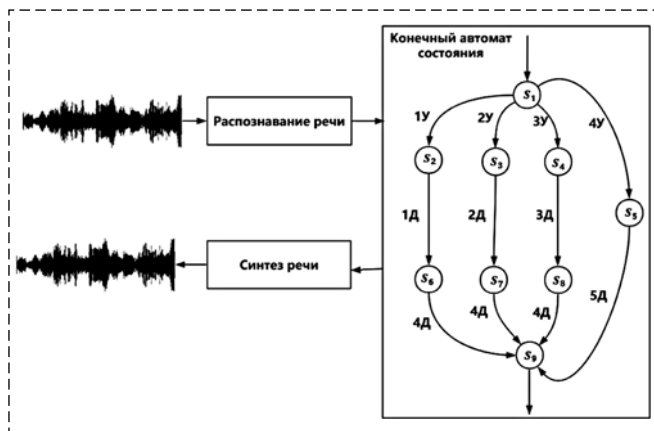


Рис. 6. Схема общего режима. Распознавание речи с помощью модели DNN-HMM; синтез речи с помощью сервиса Яндекс

Fig. 6. Scheme of the general regime. Speech recognition with DNN-HMM; speech synthesis using the Yandex service

более простой способ реализации общего режима состоит в подключении диалоговой системы к сети Интернет, содержащей необходимый сервис "Яндекс для синтеза речи" (рис. 6).

Если распознавание речевого сообщения не удастся, то система выдает сообщение: "не удалось распознавание сообщения!". Система содержит библиотеку диалогов; для управления диалогом используются такие варианты: "один ответ для нескольких запросов", "несколько ответов для одного запроса" и "ответ из сайта для специальных вопросов". Выбор варианта зависит от выполнения следующих условий и действий: 1Д (Действие) — поиск информации по условию.1; 2Д — поиск информации по условию.2; 3Д — поиск информации по условию.3; 4Д — синтез ответов; 5Д — синтез сообщения "не удалось распознавание речи"; 1У (Условие) — запросы клиента содержат функциональную информацию, например, "что ты можешь?", "что ты делаешь?"; 2У — запросы клиента содержат информацию о текущем состоянии робота; 3У — запросы клиента содержат текущую информацию о рабочем состоянии робота, о погоде и т. п.; 4У — результат распознавания речи неправильный.

Экспериментальное исследование диалога "оператор—робот"

Исследование проводили с помощью манипулятора KUKA KR3 R540. Предполагалось, что предварительно получены координаты наблюдаемых объектов с помощью открытого

исходного пакета OpenCV. Оператор способен ввести голосовую команду через микрофон РС с использованием языка Python. Диалоговая система комбинированного типа обрабатывает голосовую информацию, распознает и отправляет команды подсистеме управления манипулятором KUKA KRC4 через программное обеспечение RoboDK. В системе управления KRC4 запускается приемочный код RoboDKsynch.sh, который написан на стандартном языке управления манипулятором KUKA. Одновременно программа KUKA VARPROXY возвращает текущие обобщенные координаты (относительные углы поворота звеньев манипулятора) в компьютер для отображения текущего состояния манипулятора. Функциональная схема диалоговой системы приведена на рис. 7.

В проведенных экспериментах для распознавания речи использовалась модель DNN-HMM (см. выше), которая позволяла эффективно и быстро распознавать произвольные высказывания на русском языке с помощью языковой модели n -gram.

На рис. 8 (см. вторую сторону обложки) показана 3D-модель манипулятора KUKA KR3 R540, с помощью которой обрабатывалась диалоговая система управления реальным манипулятором, и приведен фрагмент диалога при управлении этой моделью: пользователь дает

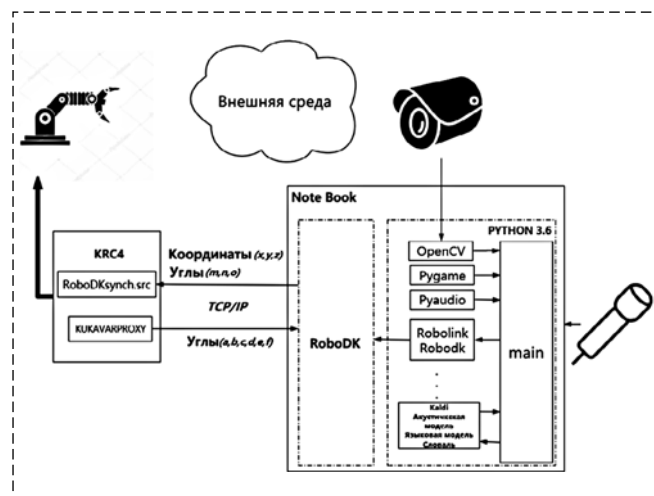


Рис. 7. Функциональная схема диалоговой системы для управления манипулятором KUKA KRC4 комбинированного типа. TCP/IP — сетевая модель передачи данных, представленных в цифровом виде; KRC4 — система управления манипулятором KUKA

Fig. 7. Functional diagram of the dialogue system for controlling the KUKA KRC4 manipulator of the combined type. TCP/IP — network model of data transmission, presented in digital form; KRC4 — KUKA manipulator control system

задание — "положить красный блок на место А"; робот после уточнения задачи отвечает: "операция выполняется", после выполнения операции, сообщается: "блок на месте". Аналогично, при управлении реальным роботом-ассистентом хирурга вводятся команды "передать скальпель", после уточнения задачи и сообщения "операция выполняется" робот осуществляет поиск нужного инструмента в соответствии с изложенной выше методикой, захватывает его и перемещает в определенное положение в рабочей области хирурга. После чего сообщает: "медицинский инструмент готов". В дальнейшем планируется добавить процедуру распознавания кисти протянутой руки хирурга и перемещение инструмента непосредственно в его руку.

Некоторые затруднения, возникавшие в процессе эксперимента, были связаны с тем, что для работы системы распознавания требуется полная тишина и отсутствие резких изменений громкости. Диалоговая система, организованная на основе теории конечных автоматов, ведет диалог с клиентом-оператором с использованием практически естественного языка, в чем состоит ее преимущество. Однако система не лишена недостатков. Система работает только по заранее определенной последовательности высказываний для диалога данного типа. Она не может прерывать текущий диалог и самостоятельно перейти к следующему диалогу, что ограничивает возможности диалога робота и оператора. Эти ограничения могут быть преодолены при дальнейших исследованиях.

Заключение

Проведенные эксперименты показали, что реализованная объектно-ориентированная система диалогового управления роботом, построенная на основе технологии конечных автоматов состояния, успешно решает задачу управления манипулятором с помощью голосовых команд, которые, в свою очередь, распознаются с использованием комбинированного метода DNN-HMM. Получены достаточно высокие точность (92 %) и скорость автоматического распознавания речи, что позволяло вести диалог в темпе, близком к темпу естественной речи.

Отметим, что чем больше слов нужно распознать, тем меньше скорость и ниже точность

распознавания. На практике глубокие нейронные сети применяются для преобразования от голосовых сигналов к слогам. Для распознавания слов и фраз нужно применять марковскую и языковую модели, хотя это и повысит как сложность, так и время обучения. Преимуществом рассмотренной выше объектно-ориентированной диалоговой системы является упрощение преобразования от голосовых сигналов к слову.

Список литературы

1. **Jurafsky D., Martin J. H.** Speech and Language Processing: An introduction to natural language processing, computational linguistics, and speech recognition. Pearson, 2014. P. 273—543.
2. **Sergienko R.** Text Classification for Spoken Dialogue Systems. Institute of Telecommunications and Institute of Artificial Intelligence, 2016. P. 17—58.
3. **Mansour A. H., Salh G. Z. A., Mohammed K. A.** Voice Recognition using Dynamic Time Warping and Mel-Frequency Cepstral Coefficients Algorithms // International Journal of Computer Applications. 2015. P. 34—41.
4. **Yu Z. S., Kobayashi H.** An Efficient Forward-Backward Algorithm for an Explicit-Duration Hidden Markov Model // IEEE Signal Processing Letters. 2003. P. 11—14.
5. **Tu S.** Derivation of Baum-Welch Algorithm for Hidden Markov Models. URL: <https://people.eecs.berkeley.edu/~stephen-tu/writeups/hmm-baum-welch-derivation.pdf>
6. **Tao C.** A generalization of discrete hidden Markov model and of Viterbi algorithm // Department of Computer Science. 1992. P. 1381—1387.
7. **Rabiner L. R.** A tutorial on hidden Markov models and selected applications in speech recognition // Proceeding of the IEEE. 1989. P. 257—286.
8. **Rabiner L., Juang B. H.** Fundamentals of Speech Recognition. Prentice-Hall, Upper Saddle River, 1993. P. 321—386.
9. **Arisoy E., Sainath T., Kingsbury B., Ramabhadran B.** Deep neural network language model // In proceedings of the Joint Human Language Technology Conference and the North American Chapter of the Association of Computational Linguistics Workshop. 2012. P. 20—28.
10. **Dong Y., Li D.** Automatic Speech Recognition (A Deep Learning Approach). London: Springer-Verlag, 2015. P. 13—48
11. **Pauls A., Klein D.** Faster and smaller N-gram language models // Annual Meeting of the Association for Computational Linguistics: Human Language Technologies. 2011. P. 258—267.
12. **Ющенко А. С.** Диалоговое управление роботами на основе нечеткой логики // Тр. Междунар. науч.-техн. конф. "Экстремальная робототехника", 25—26 сентября 2012. Санкт-Петербург: Политехника-сервис, 2012. С. 29—36.
13. **Meza-Ruiz I. V., Riedel S., Lemon O.** Spoken language understanding in dialogue systems, using a 2-layer Markov logic network // Improving semantic accuracy. Semantics and Pragmatics of Dialogue (LONDIAL'08). 2008. P. 191—192.
14. **Williams J. D.** Web-style ranking and slu combination for dialog state tracking // Annual Meeting of the Special Interest Group on Discourse and Dialogue. 2014. P. 282—291.
15. **Henderson M., Thomson B., Young S. J.** Deep Neural Network Approach for the Dialog State Tracking Challenge // Proceedings of SIGDIAL. 2013. P. 467—471.
16. **Thomson B., Young S.** Bayesian update of dialogue state: A POMDP framework for spoken dialogue systems // Computer Speech & Language. 2010. P. 562—588.

17. **Pieraccini R., Huerta M. J.** Where do we go from here? research and commercial spoken dialog systems // SIGdial Workshop on Discourse and Dialogue. 2005. P. 1–24.

18. **Thomson B., Young S.** Bayesian update of dialogue state: A POMDP (the partially observable Markov decision process) framework for spoken dialogue systems // Computer Speech & Language. 2010. P. 562–588.

19. **Yuschenko A. S., Morozov D. N., Zhonin A. A.** Speech control for mobile Robotic systems // Proc. of 4th International Conference "Mechatronic Systems and Materials" MSM-2008, Bialostok, Poland, July 2008. P. 14–17.

20. **Жонин А. А.** Алгоритм обучения менеджера диалога речевой диалоговой системы управления роботом // Интегрированные модели и мягкие вычисления в искусственном интеллекте. Сб. научн. тр. междунар. конф. М.: Физ.-мат. лит., 2011. С. 395–406.

21. **Ющенко А. С.** Интеллектуальное планирование в деятельности роботов // Мехатроника, автоматизация, управление. 2005. № 3. С. 5–18.

22. **Huang J., Rathod V., Sun C., Zhu M. I., Korattikara A.** Speed/accuracy trade-offs for modern convolutional object detectors // Computer Vision and Pattern Recognition, 2017.

Dialogue System of Controlling Robot Based on the Theory of Finite-State Automata

Yin Shuai, shuai.yin@yandex.ru, **A. S. Yuschenko**, yusch@bmstu.ru,

Bauman Moscow State Technical University, 105005, Moscow, Russian Federation

Corresponding author: **Yin Shuai**, Post-Graduate Student of Robotic Systems and Mechatronics Department, Bauman Moscow State Technical University, Moscow, 105005, Russian Federation, e-mail: scholess18@mail.ru

Accepted on July 29, 2019

Abstract

The article discusses the system of dialogue control manipulation robots. The analysis of the basic methods of automatic speech recognition, speech understanding, dialogue management, voice response synthesis in dialogue systems has been carried out. Three types of dialogue management are considered as "system initiative", "user initiative" and "combined initiative". A system of object-oriented dialog control of a robot based on the theory of finite state machines with using a deep neural network is proposed. The main difference of the proposed system lies in the separate implementation of the dialogue process and robot's actions, which is close to the pace of natural dialogue control. This method of constructing a dialogue control robot allows system to automatically correct the result of speech recognition, robot's actions based on tasks. The necessity of correcting the result of speech recognition and robot's actions may be caused by the users' accent, working environment noise or incorrect voice commands. The process of correcting speech recognition results and robot's actions consists of three stages, respectively, in a special mode and a general mode. The special mode allows users to directly control the manipulator by voice commands. The general mode extends the capabilities of users, allowing them to get additional information in real time. At the first stage, continuous speech recognition is built by using a deep neural network, taking into account the accents and speech speeds of various users. Continuous speech recognition is a real-time voice to text conversion. At the second stage, the correction of the speech recognition result by managing the dialogue based on the theory of finite automata. At the third stage, the actions of the robot are corrected depending on the operating state of the robot and the dialogue management process. In order to realize a natural dialogue between users and robots, the problem is solved in creating a small database of possible dialogues and using various training data. In the experiments, the dialogue system is used to control the KUKA manipulator (KRC4 control) to put the desired block in the specified location, implemented in the Python environment using the RoboDK software. The processes and results of experiments confirming the operability of the interactive robot control system are given. A fairly high accuracy (92 %) and an automatic speech recognition rate close to the rate of natural speech were obtained.

Keywords: deep neural network, hidden Markov model, training, dialogue control, finite state machine, robot control system

For citation:

Yin Shuai, Yuschenko A. S. Dialogue System of Controlling Robot Based on the Theory of Finite-State Automata, *Mekhatronika, Avtomatizatsiya, Upravlenie*, 2019, vol.20, no. 11, pp. 686–695.

DOI: 10.17587/mau.20.686-695

References

1. **Jurafsky D., Martin J. H.** Speech and Language Processing: An introduction to natural language processing, computational linguistics, and speech recognition, Pearson, 2014, pp. 273–543.

2. **Sergienko R.** Text Classification for Spoken Dialogue Systems, Institute of Telecommunications and Institute of Artificial Intelligence, 2016, pp. 17–58.

3. **Mansour A. H., Salh G. Z. A., Mohammed K. A.** Voice Recognition using Dynamic Time Warping and Mel-Frequency Cepstral Coefficients Algorithms, *International Journal of Computer Applications*, 2015, pp. 34–41.

4. **Yu Z. S., Kobayashi H.** An Efficient Forward-Backward Algorithm for an Explicit-Duration Hidden Markov Model, *IEEE Signal Processing Letters*, 2003, pp. 11–14.

5. **Tu S.** Derivation of Baum-Welch Algorithm for Hidden Markov Models, available at: <https://people.eecs.berkeley.edu/~stephentu/writeups/hmm-baum-welch-derivation.pdf>

6. **Tao C.** A generalization of discrete hidden Markov model and of Viterbi algorithm, Department of Computer Science, 1992, pp. 1381–1387.

7. **Rabiner L. R.** A tutorial on hidden Markov models and selected applications in speech recognition, *Proceeding of the IEEE*, 1989, pp. 257–286.

8. **Rabiner L., Juang B. H.** Fundamentals of Speech Recognition, Prentice-Hall, Upper Saddle River, 1993, pp. 321–386.
9. **Arisoy E., Sainath T., Kingsbury B., Ramabhadran B.** Deep neural network language models, *In proceedings of the Joint Human Language Technology Conference and the North American Chapter of the Association of Computational Linguistics Workshop*, 2012, pp. 20–28.
10. **Dong Y., Li D.** Automatic Speech Recognition (A Deep Learning Approach), Springer-Verlag, London, 2015, pp. 13–48.
11. **Pauls A., Klein D.** Faster and smaller N-gram language models, *Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, 2011, pp. 258–267.
12. **Yuschenko A. S.** Interactive robot control based on fuzzy logic, *Proceedings of the international scientific-technical conference "Extreme Robotics"*, September 25–26 2012, St. Petersburg, Polytechnic service, 2012, pp. 29–36 (in Russian).
13. **Meza-Ruiz I. V., Riedel S., Lemon O.** Spoken language understanding in dialogue systems, using a 2-layer Markov logic network, *Improving semantic accuracy. Semantics and Pragmatics of Dialogue (LONDIAL'08)*, 2008, pp. 191–192.
14. **Williams J. D.** Web-style ranking and slu combination for dialog state tracking, *Annual Meeting of the Special Interest Group on Discourse and Dialogue*, 2014, pp. 282–291.
15. **Henderson M., Thomson B., Young S. J.** Deep Neural Network Approach for the Dialog State Tracking Challenge, *Proceedings of SIGDIAL*, 2013, pp. 467–471.
16. **Thomson B., Young S.** Bayesian update of dialogue state: A POMDP framework for spoken dialogue systems, *Computer Speech & Language*, 2010, pp. 562–588.
17. **Pieraccini R., Huerta M. J.** Where do we go from here? research and commercial spoken dialog systems, *SIGdial Workshop on Discourse and Dialogue*, 2005, pp. 1–24.
18. **Thomson B., Young S.** Bayesian update of dialogue state: A POMDP (the partially observable Markov decision process) framework for spoken dialogue systems, *Computer Speech & Language*, 2010, pp. 562–588.
19. **Yuschenko A. S., Morozov D. N., Zhonin A. A.** Speech control for mobile Robotic systems, *Proc. of 4th International Conference "Mechatronic Systems and Materials" MSM-2008*, Byalostok, Poland, July, 2008, pp. 14–17.
20. **Zhonin A. A.** Algorithm for learning the dialogue manager of the dialogue robot control system, *Integrated models and soft computing in artificial intelligence. 8th scientific papers of the international conference*, Moscow, Phys. mat. Lit. 2011, pp. 395–406 (in Russian).
21. **Yuschenko A. S.** Intellectual planning in the activities of robots, *Mekhatronika, Avtomatizatsiya, Upravlenie*, 2005, no. 3, pp. 5–18 (in Russian).
22. **Huang J., Rathod V., Sun C., Zhu M. I., Korattikara A.** Speed/accuracy trade-offs for modern convolutional object detectors, *Computer Vision and Pattern Recognition*, 2017.



21 февраля 2020 г.
в г. Новокузнецк на базе Научно-исследовательского центра
«МашиноСтроение» (НИЦ МС) состоится

IV Международная научно-практическая конференция "МЕХАТРОНИКА, АВТОМАТИКА И РОБОТОТЕХНИКА"



Секции конференции:

1. Роботы, мехатроника и робототехнические системы.
2. Методы и техника создания и исследования интеллектуальных машин.
3. Механика и управление движением машин.
4. Механизация, автоматизация и управление технологическими процессами и производствами.
5. Методы контроля и диагностики в машиностроении.
6. Информационно-измерительные и управляющие системы.
7. Математическое и программное обеспечение вычислительных машин и комплексов.

По итогам конференции издается сборник трудов с присвоением УДК, ББК, ISSN.

Сборник индексируется в базе данных РИНЦ.

Всем публикациям постатейно присваивается цифровой идентификатор DOI.

Подробную информацию о конференции см. на сайте:
<http://srcms.ru>