

ЭКСПЕРИМЕНТАЛЬНАЯ И ПРИКЛАДНАЯ ЛИНГВИСТИКА

УДК 811.111+82-36

АЛГОРИТМЫ ВЕДЕНИЯ ЧЕЛОВЕКО-МАШИННОГО ДИАЛОГА ГУМАНОИДНЫМИ РОБОТАМИ *NAO*

О.А. Турбина, Е.А. Фостаковский

Рассмотрена актуальная тема разработки алгоритма ведения человеко-машинного диалога гуманоидными роботами. Проведен обзор и описание существующих теоретических и прикладных исследований в этой области, включая классификацию уже разработанных разными авторами алгоритмов по уровню их сложности.

Ключевые слова: алгоритм, гуманоидный робот, человеко-машинный диалог, диалоговая система.

В современном мире компьютерные технологии активно внедряются во всё новые и новые сферы человеческой жизни. В этой связи особенно актуальны разработки способов обеспечения человеко-машинного диалога вообще и общения гуманоидного робота с пользователем-человеком на естественном языке – в частности.

В настоящее время одними из наиболее популярных и доступных гуманоидных роботов являются гуманоидные роботы *Nao* (*Nao*) французской компании Альдебаран Роботикс (*Aldebaran Robotics*). Исследования по созданию систем искусственного интеллекта для таких роботов, позволяющих им становиться участниками полноценного диалога, ведутся учеными разных стран достаточно активно, о чем свидетельствуют труды таких авторов, как Г. Уилкок (*G. Wilcock*), К. Джокинен (*K. Jokinen*) (Финляндия), А. Ксапо (*A. Csapo*) (Венгрия), Д. Энэстейжеу (*D. Anastasiou*), М. Киршнер (*M. Kirschner*) (Германия) и других. Однако сегодня эта проблема пока далека от своего окончательного решения.

Данная статья посвящена краткому обзору существующих разработок в рассматриваемой области и выделению тех из них, которые представляются наиболее перспективными.

Говоря о системах, позволяющих гуманоидным роботам вести диалог на естественном языке с пользователем-человеком, прежде всего, отметим, что различают несколько уровней подобных систем. Системы первого уровня представляют собой линейную модель «вопрос пользователя – ответ робота» и функционируют по принципу запрограммированной реакции: пользователь-человек дает роботу голосовую команду из заранее заданного набора команд, в ответ на которую робот произносит один из заранее запрограммированных в его памяти текстов-ответов. В действи-

тельности, подобные системы не являются диалоговыми в строгом смысле, поскольку речь робота по сути представляет собой монолог, произнесенный в ответ на реплику пользователя. При этом каждая речевая реакция робота следует «вне контекста», ибо ответ на любую последующую реплику-команду для него не связан и, тем более, не обусловлен тем, что им говорилось и воспринималось ранее. Системы этого типа состоят из двух основных компонентов – модуля распознавания речи, определяющего, какая именно команда была отдана роботу, и модуля синтеза речи, который воспроизводит («произносит») ответные реплики робота. Подобные модели могут применяться, например, для создания простых вопросно-ответных систем в различных областях, о чем пишет финский исследователь Грэхэм Уилкок в своей статье «*ВикиТок: Открытая диалоговая система, использующая Википедию как источник знаний*» [9].

В диалоговых системах второго, более высокого уровня робот проявляет коммуникативную инициативу и задает пользователю-человеку вопросы, направленные на выполнение некоторой коммуникативной цели. Для робота коммуникативная цель в данном случае состоит в том, чтобы посредством вопросов выяснить у пользователя некоторые сведения и, условно говоря, вписать их в некую таблицу с результатами. Модель функционирования таких систем более сложная, поскольку, робот здесь запоминает ответы пользователя и использует их для «заполнения» электронных таблиц данных [9].

Например, подобным образом устроены системы автоматического бронирования транспортных билетов. По сравнению с системами предыдущего уровня, в таких моделях результаты коммуникации сохраняются в памяти робота, и могут оказывать влияние на его дальнейшее языковое

поведение. Так, в случае, если роботом не был распознан ответ пользователя на вопрос: «*Каков конечный пункт вашего путешествия?*», ему (пользователю) может быть задан роботом дополнительный вопрос, уточняющий эту информацию, например: «*Не могли бы вы повторить, каков конечный пункт вашего путешествия?*». Как уже было сказано, устройство таких систем отличается от устройства систем первого уровня в сторону усложнения. Системы второго уровня состоят из следующих основных компонентов: модуль анализа речи, модуль синтеза речи, а также конечный автомат, контролирующий непосредственно процесс коммуникации. Количество состояний автомата зависит от конкретной коммуникативной цели, которая должна быть выполнена в финале конкретного диалога. Г. Уилкок приводит в качестве примера устройство конечного автомата, задачей которого является помощь пользователю в бронировании авиабилета. Автомат начинает свою работу со стандартного для любого диалога состояния приветствия, находясь в котором, система приветствует пользователя и сразу же переходит в следующее состояние. При бронировании авиабилета, в состоянии «узнать город отправления» (`getDepartureCity`) автомат выясняет, из какого пункта пользователь намерен совершить вылет: активируется модуль синтеза речи, адресуящий пользователю соответствующий вопрос. Затем подключается модуль распознавания речи, задача которого воспринять и распознать ответ пользователя. Список возможных ответов ограничен и определен заранее, что значительно облегчает работу модуля. В данном случае это список городов, в которые возможно забронировать полет из данного аэропорта. Если ответ пользователя был распознан, то система переходит в следующее состояние, если нет – пользователю задается дополнительный уточняющий вопрос. Схожим образом, переходя от состояния к состоянию, система выясняет у пользователя город прибытия, даты вылета и прочую необходимую информацию. Предпоследним состоянием является состояние «подтвердить правильность полученной информации» (`giveSummary`), войдя в которое, система вслух проговаривает всю полученную от пользователя информацию и задает вопрос, всели указано верно. Если пользователь отвечает утвердительно, система переходит в «прощальное» состояние «попрощаться с пользователем» (`sayGoodbye`) и бронирует билет. Если же пользователь отвечает отрицательно, то система произносит соответствующую общую фразу, например: «*Что ж, в таком случае, повторим все сначала*», и снова переходит в состояние «узнать город отправления» [9].

Оба рассмотренных типа систем относятся к так называемым «закрытым диалоговым системам» (`closed-domain systems`), которые способны к ведению человеко-машинного диалога на фикси-

рованную, заранее заданную тематику. Им противопоставляются так называемые «открытые диалоговые системы» (`open-domain systems`), которые поддерживают ведение человеко-машинного диалога на множество заранее не заданных тем, что поднимает их ещё на уровень выше по сравнению с закрытыми диалоговыми системами [1, 3, 9].

Количество возможных тем для разговора в открытых диалоговых системах напрямую зависит от заложенной в систему информации. Г. Уилкок упоминает систему ВикиТок, источником знаний для которой является интернет-энциклопедия Википедия, но источником знаний для системы может служить любой соответствующим образом предобработанный гипертекстовый корпус достаточной величины. Преимуществом Википедии в этом плане является то, что она общедоступна, постоянно обновляется и дополняется, в силу чего количество возможных тем для разговора постоянно увеличивается по мере роста объема информации по уже представленным темам [1, 9].

Поскольку концепция ведения роботом диалога на свободную тематику гораздо сложнее, чем человеко-машинный диалог на заранее заданную тему, алгоритм ведения свободного диалога будет отличаться от алгоритма диалога заданного также в сторону усложнения, хотя формально он по-прежнему будет представлять собой конечный автомат. Рассмотрим подробнее строение такого алгоритма – алгоритма третьего уровня, описание которого встречается в работах разных авторов [1, 3, 9].

Система третьего уровня по-прежнему состоит из трех основных элементов: модуля распознавания речи, модуля синтеза речи и конечного автомата, контролирующего ход диалога. Как и в случае с диалогами на заданную тему, автомат начинает свою работу с состояния приветствия, после которого происходит переход в следующее состояние – состояние выбора темы. Тема разговора может быть выбрана несколькими способами. Во-первых, если пользователь разговаривает с роботом не в первый раз, система может обратиться к списку недавних тем беседы при условии, что такой список ею ведется. Во-вторых, тема может быть выбрана из используемой базы знаний случайным способом. В-третьих, могут применяться функции «популярная тема» или «`favourites`», как, например, в упомянутой системе ВикиТок [1, 3, 9]. Иными словами, система получает доступ к списку тем, которые отметил сам пользователь как наиболее интересные. Наконец, если пользователя интересует какая-то определенная тема, он может просто продиктовать её роботу, и в работу вступает модуль распознавания речи. Здесь, однако, стоит отметить, что, поскольку список возможных тем для диалога в рассматриваемых системах чрезвычайно велик, то классическое распознавание речи по ключевым словам не вполне подходит по ряду причин. В частности, если в используемой

базе знаний представлено большое количество тем, и они никак не сгруппированы, то система с высокой долей вероятности будет совершать ошибки при распознавании. Наиболее эффективным способом решения данной проблемы представляется информационный поиск внутри используемого источника знаний, который может осуществляться как по словам, так и по частям слов. Например, если пользователь произнесет слово «закон», то система будет искать все темы, в названии которых оно встречается: *первый закон термодинамики*, *закон об авторском праве* и т. д. Затем, система представит пользователю все найденные варианты для выбора темы разговора. Также может применяться поиск по частям слов, когда пользователь последовательно по буквам называет роботу часть ключевого слова, например, «р», «о», «б», «о», «т». Наконец, на основании предоставленной ему информации робот производит поиск в источнике знаний и обнаруживает, или не обнаруживает необходимую статью/статьи, при этом эффективность поиска зависит от выбора ключевого слова и от длины продиктованного пользователем отрезка этого слова. В качестве дополнительной страховки от ошибок система может попросить пользователя подтвердить, правильно ли были распознаны его слова или фразы, например: «Итак, вы хотите поговорить о роботах Нао, верно?» [1, 9].

После выбора темы разговора, система обращается к используемому источнику знаний, из которого извлекает и «зачитывает» пользователю соответствующую информацию: вступает в работу модуль синтеза речи. Количество информации, которое робот сообщает пользователю, напрямую зависит от сложности выбранной темы и оттого, насколько подробно она представлена в источнике знаний. При выборе тем, представленных объемными текстами, перед роботом встает проблема определения количества выдаваемой пользователю информации за один раз, поскольку если робот будет говорить слишком долго, то коммуникация примет форму не диалога, а монолога, и достаточно быстро наскучит пользователю. Наиболее очевидным решением данной проблемы представляется, «зачитывание» роботом текста частями – по абзацам, либо группами предложений. Хотя для этого системе дополнительно потребуется модуль анализа текста, сложность такого анализа относительно невелика, ибо он будет ограничен поиском границ абзацев и отдельных предложений. После зачитывания первого фрагмента текста автомат переходит в следующее состояние – состояние продолжения темы, и система спросит у пользователя, хочет ли он услышать ещё что-нибудь интересное по текущей теме. Если ответ человека утвердительный, автомат остается в состоянии продолжения темы, и робот зачитает следующий фрагмент текста; и так далее, пока пользователь не решит сменить тему [3, 9].

Когда пользователь решит сменить тему раз-

говора и сообщит роботу о своем решении, конечный автомат из состояния продолжения темы вернется в состояние выбора темы.

Смена темы разговора может проходить и более плавно. В упомянутой выше статье Г. Уилкока [9], используется термин «плавные смены темы» (smooth topic shifts). Автор описывает, как происходят такие «плавные переходы» между темами, на примере робота Нао, подключенного к системе ВикиТок. Известно, что статьи в Википедии соединены между собой посредством гиперссылок, которыми могут являться слова непосредственно в тексте той или иной статьи. Например, из статьи о роботах Нао, производимых французской компанией Альдебаран Роботикс, можно перейти к статье, посвященной самой этой компании, при условии, что слова Альдебаран Роботикс являются гиперссылкой. Робот Нао, использующий Википедию как источник знаний, способен делать то же самое прямо во время разговора. Алгоритм работы в этом случае следующий: пользователь сообщает системе, что хочет поговорить о роботах Нао, система находит в Википедии соответствующую статью и начинает «зачитывать» пользователю первый её фрагмент, например: «*Роботы Нао производятся французской компанией Альдебаран Роботикс*» (фраза «Альдебаран Роботикс» выделена в тексте как гиперссылка). После того, как робот закончит чтение, пользователь может: а) попросить рассказать ему ещё что-нибудь о роботах Нао, б) спросить: «Альдебаран Роботикс?» или «А что такое “Альдебаран Роботикс?”» В первом случае робот продолжит чтение статьи, во втором случае система переходит по соответствующей гиперссылке и начинает чтение статьи, посвященной компании «Альдебаран Роботикс». Основываясь на встреченных в тексте гиперссылках, робот может и сам предложить пользователю сменить тему разговора, произнеся шаблонную фразу, например: «А хотите что-нибудь узнать о ... ?», что повышает степень естественности диалога [3, 4, 7–9].

Схожий подход к проблеме смены темы диалога в процессе человеко-машинного общения описывается в статье Мануэля Киршнера (Manuel Kirschner) «Применение древовидной модели представления диалога в интерактивных системах поиска ответов на вопросы» [7].

Распознавание гиперссылки внутри текста (например, «Альдебаран Роботикс») человеком-пользователем, воспринимающим текст на слух, может осуществляться при помощи жестов робота: именно способность жестикулировать является одним из важнейших свойств, отличающих гуманоидных роботов от негуманоидных. Поэтому для решения проблемы с обозначением гиперссылок в устной речи робота, достаточно включить в его поведение мультимодальный компонент, описанный в статье К. Джокинен «Мультимодальное общение на свободную тему с роботами Нао» и

«обучить» его жестам, какими пользуются люди для того, чтобы подчеркнуть важность произносимых слов. Таким жестом может быть, например, поднятый вверх указательный палец или поворот ладони тыльной стороной вверх или вниз при согнутой в локте руке и т. д. (более подробно об основных видах жестов, которым можно научить гуманоидного робота, см. в работах Адама Кендона и Р. Миины [6, 8]). В данном случае важно добиться, чтобы жесты всегда были синхронизированы с речью, поскольку сервоприводы, приводящие в движение конечности робота, имеют скоростные ограничения. Для тех роботов, у которых конечности не предусмотрены (роботы HAO выпускаются в трех конфигурациях – с руками и ногами; с руками, но без ног; без рук и ног, но способные двигать головой), могут применяться такие жесты, как кивок или поворот головы, изменение цвета глаз и т. д. Стоит отметить, что мультимодальный компонент, кроме обозначения гиперссылки, может использоваться также для придания диалогу максимальной естественности и создания реальной модели беседы с одушевленным собеседником. Последнее не менее важно, ибо «обучение» робота естественному поведению в диалоге с человеком является одной из главных задач в области разработки алгоритма ведения человеко-машинного диалога гуманоидными роботами [2, 5, 6, 8].

Для завершения диалога пользователь имеет возможность прервать речь робота прикосновением к одному из расположенных у него на голове сенсоров.

Таким образом, на сегодняшний день существует значительное количество способов и моделей ведения человеко-машинного диалога разной степени сложности и предназначенных, соответственно, для разных целей. Наиболее перспективным направлением в этой области представляется разработка вышеупомянутых «открытых диалоговых систем», поскольку они, в силу самой своей концепции, предполагают соблюдение пользователем минимального количества условностей во время ведения диалога с роботом. Тематика такого диалога может быть свободно выбрана пользователем.

Литература / References

1. Anastasiou D., Jokinen K., Wilcock G. Evaluation of WikiTalk – User Studies of Human-Robot Interaction. *Proc. of 15th Int. Conf. on Human-Computer Interaction, HCI 2013* (Las Vegas, NV, July 2013), Part IV, v. 8007 of Lect. Notes in Comput. Sci., pp. 32–42. Springer, 2013. doi: 10.1007/978-3-642-39330-3_4
2. Csapo A., Gilmartin E., Grizou J., Han J., Meena R., Anastasiou D., Jokinen K., Wilcock G. Multimodal conversational interaction with a humanoid robot. *Proceedings of 3rd IEEE International Conference on Cognitive Infocommunications (CogInfoCom 2012)*, Kosice, 2012.
3. Jokinen K., Wilcock G. Emergent verbal behaviour in human-robot interaction. *Proceedings of 2nd International Conference on Cognitive Infocommunications (CogInfoCom 2011)*, Budapest, 2011.
4. Jokinen K., Wilcock G. Constructive interaction for talking about interesting topics. *Proceedings of Eighth International Conference on Language Resources and Evaluation (LREC 2010)*, Istanbul, 2010.
5. Jokinen K., Wilcock G. Multimodal open-domain conversations with the Nao robot. *Fourth International Workshop on Spoken Dialogue Systems (IWSDS 2012)*, Paris, 2012.
6. Kendon A. *Gesture. Visible Action as Utterance*. Cambridge, Cambridge University Press Publ., 2005.
7. Kirschner M. Applying a focus tree model of dialogue context to interactive question answering. *Proceedings of ESSLLI'07 Student Session*, Dublin, 2007.
8. Meena R., Jokinen K., Wilcock G. Integration of gestures and speech in human-robot interaction. *Proceedings of 3rd IEEE International Conference on Cognitive Infocommunications (CogInfoCom 2012)*, Kosice, 2012.
9. Wilcock G. WikiTalk: A Spoken Wikipedia-based Open-Domain Knowledge Access System. *Proceedings of the COLING-2012 Workshop on Question Answering for Complex Domains*, Mumbai, 2012. pp. 57–69.

Турбина Ольга Александровна, доктор филологических наук, профессор, зав.кафедрой общей лингвистики, Южно-Уральский государственный университет (г. Челябинск). E-mail: toa@susu.ac.ru

Фостаковский Евгений Андреевич, аспирант кафедры общей лингвистики, Южно-Уральский государственный университет (г. Челябинск). Научный руководитель – профессор О.А. Турбина. E-mail: fostakovsky9004@gmail.com

ALGORITHMS OF HUMAN-ROBOT DIALOGUE AND THEIR IMPLEMENTATION ON NAO HUMANOID ROBOTS

O.A. Turbina, South Ural State University, Chelyabinsk, Russian Federation, toa@susu.ac.ru

E.A. Fostakovsky, South Ural State University, Chelyabinsk, Russian Federation, fostakovsky9004@gmail.ru

This paper is devoted to the urgent problem of spoken human-robot dialogue. The authors give an outline of theoretical and applied research in this sphere. The outline includes the classification of existing algorithms of spoken human-robot dialogue, the criterion being their complexity.

Keywords: algorithm, humanoid robot, human-robot dialogue, dialogue system.

Olga A. Turbina, Ph.D., professor, head of the General Linguistics Department, South Ural State University (Chelyabinsk), E-mail: toa@susu.ac.ru

Evgeniy A. Fostakovsky, postgraduate student at the General Linguistics Department, South Ural State University (Chelyabinsk), scientific adviser – professor Turbina O.A. E-mail: fostakovsky9004@gmail.ru

Поступила в редакцию 20 октября 2013 г.