



ТЕХНОЛОГИИ

РОБОТЫ С СЕРДЦЕМ

Прежде чем мы сможем разделить нашу жизнь с роботами, нужно научить их понимать и имитировать чувства человека

Паскаль Фанг

«Извините,
я вас
не расслышал».

Такой может оказаться первая фраза, выражающая чувство сожаления, которую произнесет офисный робот. В конце 1990-х гг. бостонская компания *SpeechWorks International*, занимающаяся распознаванием речи и технологиями озвучивания текста, начала поставлять предпри-

ятиям по специальным заказам программные средства, которые позволяют роботам использовать эту и другие фразы. За прошедшие с тех пор годы мы привыкли разговаривать с машинами. Почти каждый телефонный звонок в службу по работе с клиентами начинается с общения с роботом. Сотни миллионов людей постоянно носят с собой персонального интеллектуального помощника. Мы можем обратиться к *Siri* или другой вопросно-ответной системе, чтобы узнать адрес ресторана, позвонить своим друзьям или найти любимую песню. Эти устройства способны имитировать и малопривлекательные особенности личности. Так, человек может спросить: «*Siri*, ты любишь меня?» — и услышать в ответ: «Я не способен любить».

ОБ АВТОРЕ

Паскаль Фанг (Pascale Fung) — профессор электроники и компьютерной инженерии Гонконгского университета науки и технологии. За вклад в налаживание взаимодействия между человеком и машинами избрана членом международного Института инженеров по электротехнике и электронике (IEEE) и членом Международной ассоциации по речевой коммуникации (ISCA).



Но машины не всегда срабатывают так, как нам бы этого хотелось. Программы распознавания и синтеза речи иногда ошибаются. Устройства часто не могут понять, чего именно мы хотим. Им недоступны эмоции и чувство юмора, сарказм и ирония. Если в будущем мы собираемся теснее взаимодействовать с машинами — будь то интеллектуальный пылесос или роботизированная человекоподобная сиделка, — понадобится, чтобы они не просто понимали слова, которые мы произносим, но чтобы они стали такими, как мы. Другими словами, они должны понимать и разделять наши чувства, т.е. уметь сопереживать.

Созданием таких устройств мы и занимаемся в возглавляемой мною лаборатории в Гонконгском университете науки и технологии. Эмпатичные (способные сопереживать) роботы могут быть очень полезны обществу. Это будут не просто помощники, а компаньоны и собеседники, дружелюбные и надежные, угадывающие наши физические и эмоциональные запросы. В ходе общения с человеком они будут обучаться и делают нашу жизнь богаче, а работу эффективнее. Они научатся извиняться за свои ошибки и спрашивать разрешения сделать то или иное, заботиться о пожилых людях и обучать наших детей. Они пожертвуют собой, чтобы спасти вашу жизнь в критической ситуации, а это проявление максимальной степени эмпатии.

Некоторые роботы, способные имитировать эмоции, уже поступили в продажу. В их числе — *Pepper*, небольшой человекоподобный робот-компаньон, разработанный французской компанией *Aldebaran Robotics* для японской фирмы *Softbank Mobile*, а также *Jibo*, шестифутовый настольный персональный робот-помощник, созданный группой инженеров, в состав которой входит Роберто Пьераччини (Roberto Pieraccini), бывший директор по диалоговым технологиям компании *SpeechWorks*. Эмпатичные роботы делают свои

первые шаги, но уже появляются инструменты и алгоритмы, которые смогут их значительно усовершенствовать.

Модуль сопереживания

Я заинтересовалась созданием эмпатичных роботов шесть лет назад, когда моя исследовательская группа разработала первый китайский аналог *Siri*. Я пришла в восторг от того, как непринужденно вели себя пользователи в общении с персональными системами-помощниками и как они расстраивались, когда машины их не понимали. Мне стало очевидно, что ключом к созданию машин, способных понимать чувства человека, могут стать алгоритмы распознавания речи, подобные тем, которым я посвятила 25 лет своей исследовательской деятельности.

По сути, любая интеллектуальная машина представляет собой набор программных модулей, каждый из которых решает определенную задачу. Один такой модуль служит для обработки человеческой речи, другой — для распознавания объектов на изображениях, полученных его бортовой видеокамерой, и т.д. У эмпатичного робота есть сердце, но это сердце — часть компьютерной программы под названием «модуль эмпатии». Для того чтобы понять чувства человека и сообщить роботу, как следует реагировать, модуль эмпатии анализирует мимику, интонации речи и само ее содержание.

Общаясь друг с другом, мы неосознанно используем различные сигналы, по которым судим об эмоциональном состоянии друг друга. Я говорю в частности о языке мимики и жестов, об интонации и, наконец, о содержании разговора. Прежде чем создавать модуль эмпатии, нужно идентифицировать те особенности человеческого общения, которые машина сможет использовать для распознавания чувств, а затем разработать соответствующие алгоритмы.

ОСНОВНЫЕ ПОЛОЖЕНИЯ

- Поскольку мы все чаще взаимодействуем с управляемыми голосом и жестами машинами, хотелось бы, чтобы они умели распознавать эмоции и понимать такие тонкие нюансы, как ирония или сарказм.
- Чтобы это стало возможно, нужно снабдить машины модулем эмпатии — программным обеспечением, которое вычленяет из речи и поведения человека специфические сигналы и соответственно им формирует реакцию робота.
- Разработка эмпатичных роботов находится в начальной стадии, но ученые уже используют технологию обработки сигналов, алгоритмы, которые обучают машины, а также средства анализа настроения для создания виртуальных роботов, способных сопереживать человеку.

Когда встала задача обучить машину реагировать на эмоции разговаривающих людей, мы решили в дополнение к пониманию значения самих слов научить ее распознавать основные акустические особенности речи, поскольку именно это характерно для людей. Мы вряд ли думаем об этом в подобных терминах, но общение между людьми — это постоянная обработка сигналов. Наш мозг распознает эмоции, передаваемые речью, выделяя звуковые сигналы, которые несут информацию о нашем состоянии: напряжении, радости, страхе, раздражении, досаде и т.д. Находясь в приподнятом настроении, мы говорим быстрее и более высоким голосом. Когда мы подавлены, наша речь ставится тусклой и монотонной. Используя технологии обработки сигналов, компьютеры могут распознать все эти оттенки, так же как полиграф распознает скачки артериального давления, пульса и электропроводности кожи. Чтобы уловить состояние стресса, мы использовали так называемое контролируемое обучение, или обучение с учителем, — и машина стала распознавать соответствующие звуковые сигналы.

Короткая запись человеческой речи может состоять всего из нескольких слов, но этого достаточно, чтобы по интонации уловить огромное количество сигналов. Мы начали с обучения машины распознаванию состояния стресса, основываясь на отрывках разговоров студентов моего университета, который они прозвали «Гонконгским университетом стресса и напряжения». Задавая студентам 12 нарастающих по уровню напряженности вопросов, мы впервые создали базу стрессовых эмоций, выражаемых на английском языке, а также на мандаринском и кантонском диалектах китайского языка. К тому времени, когда мы набрали примерно десять часов записей, наши алгоритмы могли безошибочно распознавать состояние стресса в 70% случаев, что очень близко к возможностям обычных людей.

Параллельно другие мои сотрудники обучали машину распознавать настроение, выражаемое музыкой, анализируя только акустические особенности звука (т.е. не ориентируясь на слова). Настроение, в отличие от эмоций, — это состояние души, которое сохраняется дольше, чем звучит музыка. Исследования начались с того, что мы отобрали 5 тыс. отрывков вокальных музыкальных произведений разных жанров, исполняемых на основных европейских и азиатских языках. Несколько сотен этих отрывков уже были распределены по 14 разным категориям в зависимости от создаваемого ими настроения.

Используя свои электронные средства, мы выделили примерно 1 тыс. базовых атрибутов сигнала, характерных для каждого отрывка, таких как громкость звука, основная частота, гармония и т.д., и затем использовали такую «меченую музыку» для обучения 14 различных программных

«классификаторов», каждый из которых определял, к какой категории относится ли тот или иной музыкальный отрывок. Например, один классификатор воспринимал только веселую музыку, другой — только грустную. Все 14 классификаторов работали совместно. Если «веселый» классификатор ошибочно относил грустную песню к разряду веселых, то в следующем цикле обучения он переобучался. В каждом цикле переобучался самый слабый классификатор, и вся система совершенствовалась. В результате, прослушав множество музыкальных отрывков, машина определяла, какому настроению соответствует каждый из них. Со временем она научилась различать настроение любого отрывка музыкального произведения, как это свойственно многим из нас. Отталкиваясь от этих исследований, я вместе со своими бывшими студентами основала компанию *Ivo Technologies* по созданию эмпатичных устройств для бытовых нужд. Первое из них, *Moolbox*, представляет собой домашний интеллектуальный развлекательно-информационный центр, который отвечает за характер музыки и освещение в каждой комнате, а также реагирует на эмоции хозяина.

Угадывая желания

Чтобы воспринимать и различать сарказм, иронию и другие сложные нюансы, машина должна не просто распознавать эмоции по характерным звуковым особенностям, но и понимать основной смысл разговора и сравнивать его содержание с эмоциями собеседников.

Система распознавания речи совершенствуется с использованием данных, собираемых с начала 1980-х гг., и сегодня находится на достаточно высоком уровне. Но между транскрибированием и пониманием речи существует большая разница.

Рассмотрим цепочку когнитивных, нервных и мышечных актов, сопровождающих разговор одного человека с другим: сначала человек формулирует свои мысли, затем подбирает слова и проговаривает их, а собеседник осмысливает сказанное. Речевая цепочка между человеком и машиной строится таким образом: звуковые колебания оцифровываются и преобразуются в нужные для дальнейшей работы параметры; компьютерные программы распознавания речи трансформируют эти параметры в слова, а семантический декодер (дешифратор) воссоздает из набора слов смысл фразы.

Приступая к работе по созданию эмпатичных роботов, мы понимали, что алгоритмы, подобные тем, что идентифицируют настроение пользователя по его онлайн-высказываниям, могут помочь нам в анализе проявлений эмоций в речи. Эти обучающие алгоритмы выискивают в ней ключевые слова — грусть, страх, опасение и т.д., что наводит на мысль об одиночестве говорящего. Многократное использование разговорных словосочетаний

(например, *с'топ* — «давай же») указывает на то, что песня энергичная, заводная. Мы анализируем информацию и о стиле речи: она может быть четкой и ясной либо уклончивой, с длинными паузами, ответы — развернутыми и детализированными или сухими и лаконичными.

В ходе исследования, касающегося распознавания настроения, которое несет в себе музыка, мы создали алгоритмы для поиска лиричных сигналов. Вместо того чтобы отыскивать характерные особенности в каждом музыкальном отрывке, мы взяли цепочки слов из лирических песен и поместили их в индивидуальные классификаторы, каждый из которых определяет, передает ли та или иная цепочка одно из 14 настроений. Такие цепочки слов называют *n*-граммами — группами из *n* последовательных элементов. Для классификации настроения кроме связок слов мы использовали также частеречную разметку (этап автоматической обработки текста) этих слов как части «характеристики» стихов. С помощью *n*-грамм и частеречной разметки компьютеры могут составить статистическую аппроксимацию правил грамматики любого языка; эти правила помогают системам, таким как *Siri*, распознавать речь, а программному обеспечению, такому как *Google Translate*, — переводить текст на другой язык.

Как только устройство научится понимать смысл речи, оно сможет оценивать, в каком настроении находится говорящий. Если человек говорит, вздыхая: «Я так рад, что мне придется работать все выходные», алгоритм заметит несоответствие между эмоциональными сигналами и содержанием фразы и сделает вывод, что скорее всего она была произнесена с сарказмом. Подобным же образом устройство, способное распознавать эмоции и понимать содержание речи, может сопоставить эту информацию с другими входными сигналами и выявить более сложные нюансы. Когда пользователь говорит: «Я голоден», робот может найти оптимальное решение, основываясь на том, где оба они находятся, каковы предпочтения говорящего, который сейчас час и т.д. Если робот и его владелец находятся дома и подошло время ланча, машина может ответить: «Не сделать ли вам бутерброд?». Если они находятся в дороге, ответ будет другим: «Вы хотите, чтобы я нашел ресторан?».

Супердевушка Зара

В начале 2015 г. студенты и молодые ученые из моей лаборатории начали собирать все варианты модулей распознавания речи и эмоций, чтобы воплотить их в прототип эмпатичного робота, который мы назвали «Супердевушка Зара». На сбор нужных для обучения Зары данных у нас ушли сотни часов, но сегодня программа работает на одном ноутбуке. Пока Зара — виртуальный робот, представленный на экране как персонаж анимационного фильма.

Когда вы обращаетесь к Заре, она отвечает: «Пожалуйста, подождите, пока я изучу ваше лицо»; заложенные в Зару алгоритмы анализируют полученные с помощью вебкамеры компьютера изображения, с тем чтобы определить ваш пол и расовую принадлежность. Затем Зара пытается выяснить, на каком языке вы разговариваете (Зара понимает английский язык, мандаринский диалект китайского языка и совершенствуется во французском), и задает вам несколько вопросов: «Какое ваше самое раннее воспоминание? Расскажите о вашей маме. Как вы провели свой последний отпуск? Сочините рассказ, в котором фигурируют женщина, собака и дерево». Основываясь на вашей мимике, акустических особенностях голоса и содержании вашего рассказа, Зара ответит вам с выражением участия. Через пять минут общения она попытается определить особенности вашей личности и спросит, как вы относитесь к эмпатичным машинам. Это один из способов узнать о впечатлении людей об их взаимодействии с первыми эмпатичными роботами.

Зара — прототип, но, поскольку в ее основе лежат обучающие алгоритмы, по мере общения с людьми и накопления информации она становится более «умной» и чуткой. В настоящий момент ее познания почерпнуты только из взаимодействий с сотрудниками моей лаборатории. В 2016 г. мы собираемся наделить Зару телом, чтобы она стала человекоподобным роботом.

Было бы опрометчиво утверждать, что эпоха способных к сопереживанию роботов уже настала. Мы только начинаем разработку основных инструментов, необходимых «чувствительным» роботам. И даже когда на рынке появятся потомки Зары, мы не думаем, что они будут совершенны. Если честно, я прихожу к мысли, что, сосредоточившись на улучшении точности и эффективности устройств, мы теряем суть. Важно, чтобы наши машины становились более человечными, даже если они технически несовершенны. В конце концов, так устроен и сам человек. Если мы сделаем все правильно, эмпатичные устройства не станут роботами-повелителями, чего некоторые опасаются. Они будут нашими опекунами, учителями и друзьями. ■

Перевод: С.Э. Шафрановский

ДОПОЛНИТЕЛЬНЫЕ ИСТОЧНИКИ

- Андерсон М., Андерсон С. Роботы, будьте добры! // ВМН, № 11–12, 2010.
- A Multilingual Database of Natural Stress Emotion. Xin Zuo, Tian Li and Pascale Fung. Presented at the Eighth International Conference on Language Resources and Evaluation (LREC 2012), Istanbul, May 21–27, 2012.
- Видеоролик о прототипах эмпатичных роботов см. по адресу: ScientificAmerican.com/nov2015/robots