**Behind the Mic: The Science of Talking with Computers**

GEOFFREY HINTON: We come into this world with the innate abilities to learn to interact with other sentient beings. Suppose you had to interact with other people by writing little messages to them. It'd be a real pain. And that's how we interact with computers. It's much easier just to talk to them. It's just so much easier if the computers could understand what we're saying. And for that, you need really good speech recognition.

NARRATOR: The first speech recognition system was developed by Bell Laboratories in 1952. It could only recognize numbers spoken by one person. In the 1970s, Carnegie Mellon came out with the Harpy system, which was able to recognize over 1,000 words and could recognize different pronunciations of the same word.

MALE COMPUTER VOICE: Tomato.

FEMALE COMPUTER VOICE: Tomato.

NARRATOR: Speech recognition continued in the '80s with the introduction of the hidden Markov model, which used a more mathematical approach to analyzing sound waves and led to many of the breakthroughs we have today.

JEFF DEAN: You're taking in very raw audio waveforms.

MALE SPEAKER: Like you get from a microphone on your phone or whatever.

MALE COMPUTER VOICE: Cheeseburger.

FRANCOISE BEAUFAYS: We chop it into small pieces and it tries to identify which phoneme was spoken in that last piece of speech.

GEOFFREY HINTON: So a phoneme is a kind of primitive unit for expressing words.

JEFF DEAN: And then it will want to stitch those together into likely words like Palo Alto.

RAY KURZWEIL: Speech recognition today is quite good at transcribing what you've said.

MALE SPEAKER: What's the weather like in Topeka?

ROBERTO PIERACCINI: You can talk about travels. You can talk about your contacts.

RAY KURZWEIL: Like where can I get pizza?

PHONE: Here are the listings for pizza.

RAY KURZWEIL: How tall is the Eiffel Tower?

PHONE: The Eiffel Tower is--

FRANCOISE BEAUFAYS: We've made tremendous improvements very quickly.

MALE SPEAKER: Who is the 21st President of the United States?

PHONE: Chester A. Arthur was the 21st--

MALE SPEAKER: OK, Google. Where's he from?

RAY KURZWEIL: Years ago, you had to be an engineer to interact with computers. I mean, today, everybody can interact.

ROBERTO PIERACCINI: One thing, though, that is still in the infancy is the understanding.

GEOFFREY HINTON: We need a far more sophisticated language understanding model that understands what the sentence means. And we're still a very long way from having that.

ALISON GOPNIK: Our ability to use language is one of the things that helps us have culture. It's one of the things that helps us pass on traditions from one generation to another. Figuring out about how the system of language works, even though that seems like a really easy problem, it turns out to be one that's really hard but that every baby has cracked by the time they're two years old.

FEMALE CHILD: There's two L's.

FEMALE SPEAKER: There's two L's. Yeah. E-L-L-I and then--

FEMALE CHILD: E.

FEMALE SPEAKER: E.

ROBERTO PIERACCINI: Language is extremely complex and sophisticated.

BILL BYRNE: From the semantics--

RAY KURZWEIL: Irony--

FRANCOISE BEAUFAYS: Strong accents--

MALE SPEAKER: Facial expressions--

RAY KURZWEIL: Human emotion because that's part of how we communicate.

BILL BYRNE: Humor.

RAY KURZWEIL: Do I have to be careful notto offend the dinosaur?

BILL BYRNE: Language has so many different layers, and that's why it's such a difficult problem.

GEOFFREY HINTON: At present, the human brain, and the learning algorithms in the human brain, are far, far better at things like language understanding. And they're still a lot better at pattern recognition.

BILL BYRNE: So whether or not we replicate exactly what the brain does to understand language and to understand speech, is still a question.

GEOFFREY HINTON: For many, many years, we believed that neural networks should work better than the dumb existing technology that's basically just table lookup. And then in 2009, two of my students, with a little input from me, got it working better. And the first time it just worked a little bit better, but then it was obvious that this could be developed to something that worked much better. The brain has these gazillions of neurons all computing in parallel. And all of the knowledge in the brain is in the strength of the connection between neurons. What I mean by neural net is something that's simulated on a conventional computer, but is designed to work in very, very roughly the same way as the brain. So until quite recently, people got features by hand engineering. They looked at sound waves, and they did Fourier analysis. And they tried to figure out, what features should we feed to the pattern recognition system? And the thing about neural networks is they learn their own features. And in particular, they can learn features and then they can learn features of features and then they can learn features of features of features. And that's led to a huge improvement in speech recognition.

JEFF DEAN: But you can also use them for language understanding tasks. And the way you do this is you represent words in very high dimensional spaces.

GEOFFREY HINTON: We can now deal with analogies where a word is represented as a list of numbers. So for example, if I take the list of 100 numbers that represents Paris and I subtract from it France and I add to it Italy, and if I look at the numbers I've got, the closest thing is the list of numbers that represents Rome. So by first converting words into these numbers using a neural net, you can actually do this analogical reasoning. I predict that in the next five years, it will become clear that these big deep neural networks with the new learning algorithms are going to give us much better language understanding.

ALISON GOPNIK: When we started out, we thought that things like chess or mathematics or logic, those were going to be the things that were really hard. They're not that hard. I mean, we can end up with a machine that actually can do chess as well as a grandmaster can play chess. The things that we thought were going to be easy for a computer system, like understanding language, those things have turned out to be incredibly hard.

BILL BYRNE: I can't even imagine the "we've done it" moment quite yet, just because there are so many pieces of this puzzle that are unsolved, both from a science point of view, as well as from a technical implementation point of view. There's a lot of unknowns.

ALISON GOPNIK: Those are the great revolutions. They're not just when we fiddle a little with what we already know, but when we discover something completely new and unexpected.

JEFF DEAN: I think once you kind of are in the ballpark of human level performance, that will be pretty remarkable.

**ПЕРЕВОД**

GEOFFREY HINTON: Мы входим в этот мир с врожденными способностями учиться взаимодействовать с другими живыми существами. Предположим, вам нужно будет взаимодействовать с другими людьми написав для них небольшие сообщения. Это была бы настоящая боль. И так мы взаимодействуем с компьютерами. Гораздо проще поговорить с ними. Будет намного проще, если компьютеры смогут понять, что мы говорим. И для этого им нужно действительно хорошее распознавание речи.

NARRATOR: первая система распознавания речи была разработана Bell Laboratories в 1952 году. Он мог распознавать только числа, произнесенные одним человеком. В 1970-е годы вышел Carnegie Mellon с системой Harpy, которая смогла распознать более 1000 слов и могла распознавать разные произношения того же слова.

MALE COMPUTER VOICE: Помидор.

FEMALE COMPUTER VOICE: Помидор.

NARRATOR: Распознавание речи продолжалось в 80-е годы с введением скрытой Markov модели, которая использовала более математический подход к анализу звуковых волн и привело к многим прорывам, которые мы имеем сегодня.

JEFF DEAN: Вы принимаете очень сырые звуковые сигналы.

MALE SPEAKER: Как вы получаете от микрофона на вашем телефоне или что-то еще.

MALE COMPUTER VOICE: Чизбургер.

FRANCOISE BEAUFAYS: Мы раскалываем его на мелкие кусочки, и он пытается определить, какая фонема была произнесена в этой последней части речи.

GEOFFREY HINTON: Итак, фонема - это вид примитивной единицы для выражения слов.

JEFF DEAN: И тогда он захочет сшить их вместе в вероятные слова, такие как Palo Alto.

RAY KURZWEIL: Распознавание речи сегодня довольно хорошо переписывает то, что вы сказали.

MALE SPEAKER: Какая погода, как в Topeka?

ROBERTO PIERACCINI: Вы можете говорить о путешествиях. Вы можете говорить о ваших контактах.

RAY KURZWEIL: Как, где я могу получить пиццу?

PHONE: Вот списки для пиццы.

RAY KURZWEIL: Насколько высока Эйфелева башня?

PHONE: Эйфелева башня -

FRANCOISE BEAUFAYS: Мы внесли колоссальные улучшения очень быстро.

MALE SPEAKER: Кто является 21-м президентом Соединенных Штатов?

PHONE: Честер А. Артур был 21-м -

MALE SPEAKER: Хорошо, Google. Откуда он?

RAY KURZWEIL: Несколько лет назад вы должны были быть инженером для взаимодействия с компьютерами. Я имею в виду, сегодня каждый может взаимодействовать.

ROBERTO PIERACCINI: Одна вещь, еще с младенчества, это понимание.

GEOFFREY HINTON: Нам нужен гораздо более сложный язык понимающая модель, которая понимает, что означает это предложение. И мы все еще очень далеки от этого.

ALISON GOPNIK: Наша способность использовать язык является одной из вещей, которая помогает нам иметь культуру. Это одна из вещей, которая помогает нам передавать традиции от одного поколения к другому. Выяснение того, как система языка работает, хотя это кажется очень простой проблемой, это, оказывается, тот, который действительно трудно, но каждый ребенок взломал к тому времени им два года.

FEMALE CHILD: Есть два L.

FEMALE SPEAKER: Есть два L. Да. E-L-L-I, а затем -

FEMALE CHILD: E.

FEMALE SPEAKER: E.

ROBERTO PIERACCINI: Язык чрезвычайно сложный и сложный.

BILL BYRNE: Из семантики -

RAY KURZWEIL: Ирония -

FRANCOISE BEAUFAYS: Сильные акценты -

MALE SPEAKER: Лицевые выражения -

RAY KURZWEIL: Человеческая эмоция, потому что это часть того, как мы общаемся.

BILL BYRNE: Юмор.

RAY KURZWEIL: Должен ли я быть осторожным, чтоб оскорбить динозавра?

BILL BYRNE: Язык имеет очень много разных слоев, и поэтому это такая сложная проблема.

GEOFFREY HINTON: В настоящее время человеческий мозг и обучение алгоритмы в мозге человека, намного лучше при таких вещах, как понимание языка. И они все еще намного лучше распознают образ.

BILL BYRNE: Так, будем ли мы реплицировать точно, что мозг делает, чтобы понять язык и понять речь, по-прежнему остается вопросом.

GEOFFREY HINTON: На протяжении многих-многих лет, мы полагали, что нейронные сети должны работать лучше, чем туманная существующая технология, которая в основном просто поиск таблицы. А затем в 2009 году двое моих учеников, с небольшим вкладом от меня, заставили его работать лучше. И в первый раз, когда он просто работал немного лучше, но тогда было очевидно, что это может быть разработано к чему-то, что работает намного лучше. И все знания в мозге находится в силе связи между нейронами. То, что я подразумеваю под нейронной сетью, - это что-то это моделируется на обычном компьютере, но предназначен для работы очень, очень грубо так же, как мозг. Итак, до недавнего времени люди получили функции вручную. Они смотрели на звуковые волны, и они делали анализ Фурье. И они попытались выяснить, что функции, мы должны кормить систему распознавания образов? И что касается нейронных сетей они изучают свои собственные особенности.

|  |  |
| --- | --- |
| recognition system | нейронные сети |
| analyzing sound | уровень производительности |
| tremendous improvements | анализ звука |
| neural networks | система распознавания |
| level performance | огромные улучшения |

1 задание. Сопоставить.

2 задание. Сопоставить.

|  |  |
| --- | --- |
| phoneme | is section of linguistics, studying the semantic meaning of units of language |
| semantics | is a kind of primitive unit for expressing words. |
| speech recognition | is the name of technologies responsible for storing, transferring, processing, protecting and reproducing information using computers |
| computer technologies | ability to speak, speak |
| speech | is the process of converting a speech signal into digital information |

3 задание. Сопоставить.

|  |  |
| --- | --- |
| pattern | шахматы |
| chess | реализация |
| recognition | распознавание |
| implementation | представление |
| performance | шаблон |

4 задание. Записать предложения в правильной последовательности.

1. And the first time it just worked a little bit better, but then it was obvious that this could be developed to something that worked much better.
2. For many, many years, we believed that neural networks should work better than the dumb existing technology that's basically just table lookup.
3. And that's led to a huge improvement in speech recognition.
4. What I mean by neural net is something that's simulated on a conventional computer, but is designed to work in very, very roughly the same way as the brain.
5. And all of the knowledge in the brain is in the strength of the connection between neurons.
6. They looked at sound waves, and they did Fourier analysis.
7. So until quite recently, people got features by hand engineering.
8. And they tried to figure out, what features should we feed to the pattern recognition system?
9. And the thing about neural networks is they learn their own features.
10. And then in 2009, two of my students, with a little input from me, got it working better.

5 задание. Вопрос – ответ.

|  |  |
| --- | --- |
| which year was the first speech recognition system developed? | it could only recognize numbers spoken by one person |
| what could it recognize? | 1952 |
| which year Carnegie Mellon came from with the Harpy system? | 1980 |
| what could the Harpy system recognize? | 1970 |
| when the system of Markov appeared? | to recognize over 1,000  words and could recognize different pronunciations  of the same word |

6 задание. True or False

|  |  |
| --- | --- |
| in 1970 there was a system that could only recognize numbers | True / False |
| Language has so little different layers, and that's why it's such a difficult problem | True / False |
| neural net is something  that's simulated on a conventional computer,  but is designed to work in very, very roughly  the same way as the brain | True / False |
| A machine was created that could play chess as well as a grandmaster | True / False |

7 задание. Пропуски.

We can now deal with \_\_\_\_\_ where a word is represented as a \_\_\_\_\_ of numbers. So for example, if I take the list of 100 \_\_\_\_\_ that represents Paris and I subtract from it France and I add to it Italy, and if I look at the numbers I've got, the closest thing is the list of numbers that represents Rome. So by first converting words into these numbers using a \_\_\_\_\_, you can actually do this analogical reasoning. I predict that in the next five years, it will become clear that these big deep neural networks with the new learning \_\_\_\_\_ are going to give us much better \_\_\_\_\_ understanding.

8 задание. Summary

We come into this world with the innate abilities to learn to interact with other sentient beings. It's just so much easier if the computers could understand what we're saying. And for that, you need really good speech recognition.

The first speech recognition system was developed by Bell Laboratories in 1952. It could only recognize numbers spoken by one person. In the 1970s, Carnegie Mellon came out with the Harpy system, which was able to recognize over 1,000 words and could recognize different pronunciations of the same word.

Speech recognition continued in the '80s with the introduction of the hidden Markov model, which used a more mathematical approach to analyzing sound waves and led to many of the breakthroughs we have today.

We chop it into small pieces and it tries to identify which phoneme was spoken in that last piece of speech. And then it will want to stitch those together into likely words like Palo Alto.

And then in 2009, two of my students, with a little input from me, got it working better. The brain has these gazillions of neurons all computing in parallel. And all of the knowledge in the brain is in the strength of the connection between neurons. What I mean by neural net is something that's simulated on a conventional computer, but is designed to work in very, very roughly the same way as the brain. So until quite recently, people got features by hand engineering. They looked at sound waves, and they did Fourier analysis. And they tried to figure out, what features should we feed to the pattern recognition system? And the thing about neural networks is they learn their own features. And that's led to a huge improvement in speech recognition.

We can now deal with analogies where a word is represented as a list of numbers.

Those are the great revolutions. They're not just when we fiddle a little with what we already know, but when we discover something completely new and unexpected.