3 impasse de la Gare
34260 St Etienne Estréchoux France
info@transformationsociety.net
+33-6 72 40 37 60

Google Duplex

A Conversation Between Ray Gallon and Andy McDonald

Google Duplex is a new technology for conducting natural conversations to carry out "real world" tasks over the phone, such as scheduling certain types of appointments. The system makes the conversational experience as natural as possible, allowing people to speak normally, as they would to another person, without having to adapt to a machine.

Google produced demonstrations of this technology at its IO conferences in 2018 and in 2019, it extended this to making appointments on the web. It is still considered to be under development.

Inside Duplex

> *At the core of Duplex is a recurrent neural network (RNN) designed to cope with these challenges, built using Google's TensorFlow Extended (TFX) [machine learning platform]. To obtain its high precision, we trained Duplex's RNN on a corpus of anonymized phone conversation data. The network uses the output of Google's automatic speech recognition (ASR) technology, as well as features from the audio, the history of the conversation, the parameters of the conversation (e.g. the desired service for an appointment, or the current time of day) and more. We trained our understanding model separately for each task, but leveraged the shared corpus across tasks.*
>
> – ***blog post by*** *Yaniv Leviathan, Principal Engineer and Yossi Matias, Vice President, Engineering, Google.*
> *https://ai.googleblog.com/2018/05/Duplex-ai-system-for-natural-conversation.html*

Duplex must be trained to each task and each domain under human supervision. Once trained, it can carry out its tasks fully autonomously, according to Google. It also performs self-monitoring, to detect when an appointment scheduling situation is too complex for it to handle. In that case, it turns the situation over to a human.

AM: Assuming that everything Google says about Duplex is true – In the most recent demo, it made the call, it did inflections, and it understood an accent.

RG: all of which are AI features, but how does this fit into Info 4.0. Do we have a definition from Google of Duplex?

AM: Yes, we do, it's "an RNN – a neural network – on the corpus of anonymized phone conversation data." That means that it has different forms of conversation stored.

RG: That's molecular information, right?

AM: It seems to be using molecules to mix and match what the conversation is.

RG: I don't know if we can answer, but is it only using strings it has stored in its library or can it parse together new strings to deal with unexpected situations?

AM: What Google says is if there's an unexpected situation, it hands over to a human. It uses Google's automatic speech recognition engine, and deep learning. It's DeepMind, behind it. It "incorporates disfluencies" – like "mm hmm" or "uh huh" – to make it sound more natural.

RG: If you go to the classic functionalities of communication developed by Roman Jakobson in the twenties it would be the "phatic" function, acknowledgement that you're still there, and attentive. And filling time while processing your answer.

AM: Yes. And it should be able to interpret the same things from the other side – so if the person on the other end says, "uh huh," it'll wait.

RG: Or if it's in response to a question, it should be able to read that as an affirmative answer.

AM: That's right, but we don't have a lot of examples, so what we can say is it's pulling from a whole pool of snippets of conversation to be able to recognize what people are saying, and to formulate the questions properly. In that sense it could be assimilated to putting molecules into play to make the conversation.

RG: I think we can accept that it's got molecular content, but how does it decide to put that content together, to what extent is it context determined, or updated on the fly?

AM: Well it seems to be updated on the fly, as it has phrases like "I want an appointment next Friday" – so it knows in the context that's in four days' time, and adapts its response.

RG: So that would be automatically determined. The updating of that information comes from the AI system, not from a human updating the content.

AM: but the kickoff is a human asking Google, "can you book me a table for next Friday…"

RG: So we can say, from a user point of view, that it's the same type of technology as Alexa, Siri, or OK Google, except that for the first time we're asking it to be proactive in the human world, and not just in an informatic world.

AM: It will engage with the person on the other end and it will come back with an appointment, which it probably will insert into your calendar. I think it is a first example of automated 4.0. But it's not able to do anything else but fix an appointment.

RG: Also, we're not looking at an offer of information, it's a transaction.

AM: It doesn't ask for information, like what's on the menu. It's just booking a date and time.

RG: So what's the difference between a transaction and other types of information?

AM: There's less variance. In the haircut demo they ask for price, for what type of service, there are various axes, but it's one transaction after another. It's a series of yeses and nos.

RG: Do you think this could be done by a rule-based chatbot?

AM: If Google is doing it this way it's because chatbots haven't gotten around to it yet. I don't think chatbots have the voice aspect. They haven't mastered all of the language part. A chatbot can do this textually, I've no illusion about it.

RG: Would a chatbot be able to understand all the responses it would get?

AM: No, that's where Duplex is different, it can apparently interpret pretty well what the person is saying.

RG: And it's using some combination of deep learning and natural language processing to do it.

AM: It's using deep neural networks, which a chatbot can't get at

RG: There's no reason why a chatbot couldn't get at it, but that would make it very expensive

AM: Yes, the hidden thing about Duplex is we don't know what the cost is.

RG: Have they put it in service?

AM: It's in test phase in 48 states and New Zealand.

RG: But the test is in a service situation?

AM: Yes

RG: Is it reasonable, or profitable to spend so much computing power and investment to do something that you can do in five minutes, or you can ask someone else to do?

AM: My natural response to that is no, people will still make a phone call.

RG: So why are they developing this? They must have some other use in mind.

AM: In all of the literature there's no indication of what a next step would be.

RG: Let's think about what this kind of technology would enable in a business situation, could it function as a preliminary sales agent to collect information before passing it on to a human agent?

AM: There's no example of Duplex getting information back.

RG: It gets the appointment information.

AM: But it doesn't go beyond that. Where I would see this expanded to, is some small level of education – step-by-step education that would involve getting other forms of written or interactive information about a subject

RG: About a subject or about the student's work?

AM: Both. Two-way conversation. You could imagine this being used to test the knowledge of a student.

RG: There's a lot of talk now in education circles about AI having a number of interesting functions for personalizing education. One of them is ability to present material according to the person's learning styles, preferences, and possibly even current emotional state. And that goes even to changing color temperature of the lighting, or temperature in the room – fairly serious environmental regulation, besides affecting the content.

AM: One way Duplex could evolve is to expand beyond the rendezvous towards a conversation with a purpose, which in this case is either testing a student's knowledge or teaching the student.

RG: Right, and another aspect of this AI personalized learning is that the AI would be able to pick out areas where the student is having trouble and be able to provide exercises that reinforce the student's capacity in that particular area

AM: but behind that you have to imagine a huge corpus of content that the system would have to rely on

RG: presumably that would be subject-specific.

AM: The creation of that information is something we would have to train people for

RG: That touches on updating on the fly, as well, since information changes so rapidly.

AM: That's all about structured information, which is part of the business we're in. But what changes is that some people would like to learn by looking at a video or by listening. For the people who want to listen, this is one step in that direction, if Google allows it to go beyond just selling a smart agenda to its users.

RG: If you look at the typical cycle of teaching material today, schoolbooks are prepared one to two years in advance, and get into the hands of students another year later. And for certain subject matter the advance of knowledge is just so much faster than that, that textbooks, even in electronic form on a tablet, are just not going to cut it, because you can get information on the internet that is way more up to date than what's in your scholarly manual.

AM: But remember, Google has mapped all that

RG: Yes, but most of that data, at least for the moment, is considered to be privately held and proprietary, so if a small business wants to rent AI power, will they have access to all this training data, or will that be reserved for large companies with big budgets and deep pockets? Will we have a situation of information inequality?

AM: We are going to have information inequality in that the people who can't have these technologies are going to miss out anyway.

RG: I'm of the belief that these technologies are going to be prerequisites for just about everything. What we call now the digital gap between countries will disappear, because poorer countries won't have the choice, they'll have to have this infrastructure. The question becomes, if they can't afford it, who puts it in, who owns it, and who controls it?

AM: I think that's part of the deployment of 5G technologies – it should make this all more feasible, at least in theory. The thing we need to worry about is how is it going to be put together in ways that people can use, and without paying for it – students don't have a lot of money to spend. If we're moving toward a system where if you can pay the fees you get an education, well that's not a model that I like at all.

RG: If we now have a computer that can call a human and the human on the other end doesn't know if it's a machine or a human, is that ethical? Should the computer identify itself as such?

AM: Google says Duplex will identify itself as a robot

RG: So what is the advantage of having a robot that imitates a human telling you it's a robot?

AM: The only advantage is simultaneity of transactions. You can get so much done at the same time. Because people don't naturally want to talk to a robot but if it's part of the accepted system

of getting information that you know, in the final outcome, is valuable and validated, then you might do that. So that brings into play, what is the future social contract on education, right?

Duplex today doesn't solve anything; it's not brining anything very new to the table except you can ask your phone to make an appointment for you. In terms of Siri or Alexa it's just one step further. And what we've trouble finding is what is Google's agenda, where do they want to take this technology? They will be selling it as part of DeepMind, and it's up to innovators to have enough funds to experiment with, and that's where we imagine that education solutions could come from, but education doesn't have a penny.

RG: Well, governments do, and they are responsible for education. But governments are likely to think of this as a replacement for teachers, which it isn't. In fact, teachers have better things to do than spoon feed content to students. Teachers should be guiding students, for example, in how do you tell if something is fake news, how to interpret and analyze what you find on the web, and sharing experiences so they don't get caught up in their own digital bubble.

AM: I think that this kind of technology could free up another space for teachers in the supervision of education, because they don't have enough time, today, on a daily basis to step back and ask questions about what they're teaching people and where they're going.

RG: One of the odd things about our current paradigm is that curriculum is based around feeding a quantity of material to a group and doesn't take individual development into consideration.

AM: This is where this technology is potentially different, but there's a lot to develop beforehand, because you need to be able to assess the level of a student, to add, layer by layer, what's missing. The whole question of where the knowledge comes from, who's writing it, how is it formulated, that's all of the jobs that we do today, that have to be reformatted, to cater for this. That's why the real power of molecular information has to be taught.

The whole risk about Duplex is that it can be hijacked to spam – for example, to fill up a restaurant's reservations where people don't turn up. For the moment, it doesn't seem to be collecting personal data, except from the user who has Google Assistant on his phone, and Google already knows who he is. And Google will know what kind of restaurant you prefer, what clothes you buy, etc. One of the bottom lines of this discussion is we are moving to a more highly connected world, and this is just one element of it. But if I have to pay to book a restaurant, I'm not going to do that, I'm going to call the restaurant.

RG: I'm sure they have other things in mind for it. But going back to being evermore interconnected, it's a great example of what we've been calling hybrid communications between humans and machines, where we don't notice or care if it's a human or a machine we're communicating with. It's not that we don't know it's a machine, we just don't care. I think we'll be functioning more and more in this mode.

AM: What'll happen if when Duplex makes a call, there's a robot on the other end? Because the corollary to this is a robot taking appointments, right? There's no intellectual reason it shouldn't work the other way 'round.

Imagine what could happen if Duplex remembered previous conversations, the person, the intent. Imagine this in the context of education. So the conversation could be: Do you want to pick up where we left off, or review a previous subject?

RG: Yes, that's part of what we mean by personalized learning – the AI knows whom it's dealing with and adjusts its parameters accordingly.

AM: If it identifies the person it could know a lot of context elements, such as grades, areas of weakness, areas of interest, and propose adapted content.

RG: That's the expectation. Where the teacher comes in is to integrate the learning acquired through the machine, and to socialize learning by facilitating exchange amongst different students. By exchanging, students have to leave their comfort zone, and the danger of a "digital learning bubble," where they only see the things that reinforce them, and start to acquire a sense of community, of the common good – very important if we're to achieve the 17 Sustainable Development Goals (SDG's) adopted by the UN for 2030. But we're still a ways away from having all that capacity, despite great strides forward.

AM: That's right, Duplex as it is today is just the beginning of a 4.0 AI Application; it's an example of how we would need to change the way we produce content, because if you want to push content through Duplex in the future, I mean information that's valid and validated, the whole validation process has to be set up. That's a bit of a nightmare, because it's getting people to agree on how it's done, and there's no body or entity out there that has the infrastructure to organize it, and I don't think we want Google doing it.

Links

A Google Blog post that describes the system in detail and provides several demos can be found at https://ai.googleblog.com/2018/05/Duplex-ai-system-for-natural-conversation.html.

Demo from Google I/O in May, 2018: https://www.youtube.com/watch?v=D5VN56jQMWM

Sound is a bit fuzzy, but this video is interesting because it's from the point of view of the restaurant receiving the call: https://www.youtube.com/watch?v=x_FuvIwSxT4`

Live test of how this works from a user perspective, from Google Assistant. https://www.youtube.com/watch?v=DGwALqd1YxQ

Video showing how Duplex capacity is extended to on line web reservations: https://www.youtube.com/watch?v=JbkoQGMf5DI