

Natural language
processing in action

Intelligent agents Tomorrow's digital valets



By Bo Begole, PhD, UP,
Global Head of Media Technologies Lab, Huawei

What is natural language processing (NLP)?

With origins dating back to the 1950s, NLP combines AI, computer science, and computational linguistics in a way that aims to let humans and computers communicate seamlessly through language. Although conversational

agents (think HAL 9000, C3PO et al) are what most people associate with the NLP lexicon, NLP does in fact cover a diverse range of applications. A few examples include machine translation, automatic summarization (for

example, news articles), sentiment analysis (includes detecting judgments, opinions, and emotions), information extraction (such as detecting events in an email and placing it on your calendar), and parsing (understanding syntax).

Where are we now with NLP? Simple speech-based systems that understand natural language are already widely in use. AI can answer questions about things like flight times, give directions, tell you where restaurants are, and perform simple financial transactions. Such systems don't need to understand a sentence but they do need to recognize keywords that indicate the user's intention (for example, "Make a reservation ...") and parameters of the task ("... for

6pm at the French bistro.").

More advanced systems can summarize news articles and recognize complex language structures. Such systems must have a coarse understanding to compress the articles without losing the key meaning.

How does NLP work? NLP employs two main techniques: symbolic and statistical. Symbolic relies on a series of pre-programmed rules that cover grammar, syntax, and so on. Statistical uses machine learning algorithms.

Main challenges: context and ambiguity

Context

Example: "clear" can be a verb or an adjective. In this case, a machine can understand the form a word takes in a sentence through Part-of-Speech (PoS) tagging.

Sentence: *James cleared the path*

Rule: A noun is a verb if the previous tag is a pronoun (in this case "James")

Therefore, the machine knows



“clear” is a verb in the example sentence, and can work out that “path” is a noun.

Ambiguity

Word Sense Disambiguation (WSD) is used in cases of polysemy (one word has multiple meanings) and synonymy (different words have similar meanings).

Example of a polyseme: “Fix”
He fixed dinner yesterday (made)
He fixed the car yesterday (repair)

In this case, PoS tagging and syntax will yield the same result.

So, a deeper approach is required that can pinpoint exact meaning based on real-world understanding. In the previous example, it’s understanding that you can’t “repair” dinner. For WSD, WordNet is the go-to resource as the most comprehensive lexical database for the English language.

Listening is not the same as hearing

Speech interaction will be increasingly necessary as we create more devices without

keyboards such as wearables, robots, AR/VR displays, autonomous cars, and Internet of Things (IoT) devices. This will require something more robust than the scripted pseudo-intelligence that digital assistants offer today. We’ll need digital attendants that speak, listen, explain, adapt, and understand context – *intelligent agents*. Not long ago speech recognition was so bad that we were surprised when it worked at all, but now it’s so good that we’re surprised when it doesn’t work. Over the last five years, speech recognition has improved at an annual rate of 15 to 20 percent, and is approaching the accuracy at which humans recognize speech. There are three primary drivers at work here.

First, teaching a computer to understand speech requires sample data and the amount of sample data has increased 100-fold as mined search engine data is increasingly the source.

Second, new algorithms have been developed called deep neural networks that are particularly well-suited for recognizing patterns in ways that emulate the human brain.

Finally, recognition technologies have moved off of a single device to the cloud, where large data sets can be maintained, and computing cores and memory are near infinite. And though sending speech over a network

Among enthusiasts, an intelligent agent is an artificial intelligence (AI) capable of making decisions based on prior experiences. Among consumers, an intelligent agent would need a few more qualities.

may delay response, latencies in mobile networks are decreasing.

The results? My kids are increasingly talking to their smartphones, using digital assistants to request directions, ask for information, find a TV show to watch, and send messages to friends.

Speaking does not make you intelligent

But to make interaction truly natural, machines must make sense of speech as well. Today's digital assistants seem amazingly intelligent, but they actually use a superficial form of understanding called intents and mentions, which detects what task the user is trying to accomplish (intent) and the properties of the task (mentions).

Basically, the system recognizes a command phrase (usually a verb) that identifies a task domain like "call", "set an alarm for", or "find". If it doesn't find all the necessary information in the user's statement, it can ask for more details in a kind of scripted dialog.

Such assistants take commands well, but they're a far cry from a personal concierge who intuitively understands your desires and can even suggest things you

wouldn't think to ask for. Today's assistants can't go off-script when recognition fails. They often can't explain their own suggestions. They can't anticipate problems and suggest alternatives. They rarely take the initiative.

You have to spell everything out to a digital assistant, and even then you may not get what you want. Soon, we'll stop being amazed by their mimicry of intelligence and start demanding actual intelligence.

What would this look like? What would a truly intelligent agent be, that can actually converse and function as the primary interface in the coming IoT revolution as we head into the stage of Augmented Innovation, where wearables, autonomous vehicles, robots, and embedded appliances will abound?

Intelligent agents can speak, listen, and hear

Among enthusiasts, an intelligent agent is an artificial intelligence (AI) capable of making decisions based on prior experiences. Among consumers, an intelligent agent would need a few more qualities.

Conversational: Language understanding needs to be less superficial than what we have today. Computers can easily miss intent or become confused and fall

back to simple web searches. That's because the system doesn't really understand what you're saying. If it doesn't recognize the type of task it's being asked to do, it doesn't have a predefined script with which to ask for more details. A human would be able to remedy the specific misunderstanding by saying, "I'm sorry. What kind of restaurant were you looking for?"

Explanatory: With a deeper language model, a conversational system can explain why it recommends a particular action or why it thinks something is true, just as a human can, unlike the "black box" recommendation systems today. For example, if I ask my TV for a legal drama and the system recommends the Marvel show *Daredevil*, I might need an explanation at first because I might not know that the title character is a lawyer by day when he isn't cleaning up the streets with his fists at night.

Resourceful: Human assistants are resourceful. When we detect a problem, we can plan around it and suggest alternatives. A deeply intelligent agent must proactively notice, for example, that the restaurant I scheduled for lunch with a colleague is closed that day for a religious holiday (this just happened to me).

Attentive: Intelligent agents should be constantly attentive. If one of my kids tells me he just used up the milk, the agent should notice and add it to my online shopping cart without me having to tell it to.

Sociable: Intelligent agents should be aware of my engagement with other people in my environment and know when and where not to interrupt.

Context-aware: Social intelligence is actually a subset (but important enough to call out separately) of the broader category of contextual intelligence, which requires understanding the situation a person is in and proactively selecting services that he or she has used in similar situations. Near the end of dinner at a restaurant, the intelligent agent should offer to call a taxi.

Engaging: Perhaps most importantly, I want my intelligent agent to engage me and express understanding of the importance of my requests. In human conversation, a tone of urgency is met with responsiveness. Humor is met with amusement. Concern is met with suggestions. I'm not looking for a mechanical personality to replace human companionship, just a genuine conversationalist that offers a level of engagement that indicates it understands and will act on my desire for urgency, mirth, or resolution.

We'll need intelligent agents soon

Deep intelligence will be more important for tomorrow's environments than today's smartphones, because robots, autonomous cars, and smart homes will need to converse, explain, re-plan, and engage in ways appropriate to the user and situation. The same deep learning technologies that have made speech recognition surprisingly accurate can achieve this.

With so much potential at our fingertips, what else would you want to see from intelligent agents? 