Lauren Dillard

NMDM 5321: Big Data and The Media

Professor Robert Berkman

20 November 2017

Big Data and Voice Experiences

Introduction

Conceptions of futuristic computer controls in science fiction likely emerged from a need to include the viewer in the interaction. In television shows like *Star Trek* and movies like *2001: A Space Odyssey*, *Knight Rider* and *Iron Man*, protagonists receive mission-critical data, control vehicles or weapons, and, generally, advance the plot by using their voice to interact with computer systems. By giving the computer system a voice, it has been personified. It becomes another character, an artificial intelligence and sometimes-antagonist, with lines to read and occasionally a little bit of acting to do.

According to Nathan Shedroff, author of "Make It So: Interaction Design Lessons from Science Fiction," "Sci-fi (and all fiction) has an effect on culture. It exerts influence on our expectations and our aspirations and, as such, has an impact on what we think is appropriate and what we desire. … It can be a great medium to prototype and explore human issues before they become critical in real products" (Houston). Inspired by science fiction, technology companies are pushing ever forward to develop intelligent "digital personal assistants" such as Microsoft's Cortana, Amazon's Alexa, Apple's Siri and the Google Assistant. Creating voice interfaces as robust as *Star Trek's* computer or *Iron Man*'s Jarvis is largely a big data challenge; a challenge that will require mastery of natural language processing, machine learning and artificial intelligence.

History

The earliest recorded experiment that mixed computers and language was conducted in 1948. Andrew Booth at Birkbeck College at the University of London developed a dictionary lookup system that could be used to aid human translation of content from one language to another. Booth had served as a cryptanalyst during World War I and this experience informed his research. Given a typed input, the system could return a translation (Terras 123). As early as 1952, engineers at Bell Labs ventured into voice and developed AUDREY. AUDREY, the "automatic digit recognizer" was the first system that could interpret any form of voice input — just the digits 0 through 9 (Pieraccini 3).

In 1954 — an era of punch cards and batch processing — IBM partnered with Georgetown University to advance machine translation. Using typed input, the team demonstrated automatic translation of more than 60 Russian sentences into English. At the time, this was a novel use of the IBM 701, a computer traditionally use for "solving problems in nuclear physics, rocket trajectories, weather forecasting and other mathematical wizardry" ("701 Translator"). Informed by the 1957 publication of "Syntactic Structures" by Noam Chomsky, researchers working on machine translation viewed each language as a code to crack. The code comprised a lexicon (set of words) and grammar (rules for use of those words) that could form a database for translation (Bellos). The translations required high-touch editing to make them comprehensible. Israeli philosopher and mathematician Yehoshua Bar-Hillel "concluded that fully-automatic high-quality translation is impossible without knowledge" (Hancox).

The introduction of that knowledge marked the second phase in the evolution of natural language processing. As early as 1966, researchers introduced natural language processing systems to controlled data or a knowledge base called a semantic net from which they could draw inferences while interpreting language input (Jones 4). Until the late 1960s, work in artificial intelligence and machine translation was largely conducted independently. In 1969, Roger Schank of Yale University introduced Conceptual Dependency Theory. This theory created a relationship between rote machine translation and a body of knowledge. Schank proposed that "there was a predetermined set of possible relationships that made up an interlingual meaning structure. These relationships could be used either to predict conceptual items that were implicit in a sentence or, coupled with syntactic rules, to inform a parser what was missing from a meaning and where it might be found in a sentence" (Schank 245). Schank's rules added information to the computer model to make the meaning clear. Though there are many ways to construct a sentence, these rules could be used to form a single logical, coded diagram of their meaning (Schank 246).

Early attempts to master natural language processing through domain knowledge coalesced through the 1970s. SHRDLU was developed by Terry Winograd at Stanford and enabled the manipulation of virtual blocks through natural language teletype input (Winograd). First described in a paper by G.G. Hendrix in 1978, the LIFER/LADDER system "used semantic grammar to parse questions and query a distributed database" with data about US Navy ships (Rao 220).

While researchers were continuing to make advancements in natural language processing (typed) through the 1960s and '70s, breakthroughs in machine translation and spoken language recognition had stalled. It was as early as 1950 that probabilistic models of linguistics began to emerge, but it was remarks by Chomsky and others that had quashed interest in the field (Manning "Probabalistic Models"). By 1980, computing power and storage had increased dramatically per Moore's Law and researchers were returning to probabilistic modeling as a solution. By feeding loads of sample language (transcripts, books, manuscripts) into computers preparing for natural language processing tasks, researchers were enabling computers to learn just as humans do — through repetition. Rather than relying on logic written to represent language of the past, probabilistic modeling was able to account for the "variation of languages across speech communities and across time" (Manning "Probabalistic Syntax…"). Rather than relying on a single source of knowledge, machine learning would enable these NLP systems to continue to evolve over time without the need for a software update.

"We thought it was wrong to ask a machine to emulate people. After all, if a machine has to move, it does it with wheels—not by walking. If a machine has to fly, it does so as an airplane does—

not by flapping its wings. Rather than exhaustively studying how people listen to and understand speech, we wanted to find the natural way for the machine to do it," said Fred Jelinek, distinguished professor at Cornell in information theory and researcher at IBM.

By the mid 1980s, Jelinek and his team had developed an experimental transcriptions system with a 20,000 word vocabulary called Tangora ("Pioneering Speech Recognition").

By 1990, Dragon Technologies released Dragon Dictate for DOS. Dragon Dictate required slow, spaced enunciation and was priced at $9,000 (Pinola). By 1992, IBM delivered the first speech recognition product to be packaged with new computers and, by 1996, the first "large vocabulary" continuous speech recognition product called IBM MedSpeak ("Pioneering Speech Recognition"). In 1996, BellSouth released the first voice-activated telephonic system, called VAL. By the end of the decade, speech recognition had plateaued. Systems could process language as quickly as it could be spoken and offered about 80 percent accuracy. Telephonic interactive voice response systems like VAL became frustratingly ubiquitous in call centers and Dragon had released the "much-improved" Dragon Naturally Speaking for just $695 (Pinola). IBM exited the consumer market and opted to focus on voice technology designed for call centers and car navigation ("Pioneering Speech Recognition").

Simultaneously, conceptions of machine learning established in the 1950s began to shift. Using corpora of domain knowledge (databases or raw text about a given subject) paired with newly developed algorithms and increased computing power, computers could become experts. In 1989, graduate students at Carnegie Mellon University developed a computer, dubbed Deep Thought, that beat a grandmaster at chess in a regular tournament game. Later that year, chess champion Garry Kasparov defeated Deep Thought handily in two games. IBM brought the team from CMU onboard to develop an early version of Deep Blue. Thought Deep Blue lost to Kasparov in 1996, improvements to the system led to a victory in 1997 (Greenemeier).

Since the early 2000s, redoubling of computer processing power and further refinements in models of machine learning and speech recognition have been limited only by the quality of data they can leverage — including massive recorded speech corpora and written language databases — and the speed at which they can process it. “The linguistics of the twentieth century has been the linguistics of scarcity of evidence,” wrote Professor John M. Sinclair in the 1997 publication “Teaching and Language Corpora.” “The gathering and processing of evidence has until recently been limited by the discrimination and attention-span of the researchers. Gradually, some instrumentation has aided the direct study of speech, and the invention of the wire recorder and tape recorder has been making a big contribution to the study of spoken word since the 1950s. But the quantity of data has been chronically inadequate for any reliable statements about grammar, vocabulary, usage, semantics or pragmatics” (Wichmann 27).

By early 2000, web crawlers and information extraction had become ubiquitous on the web. The extracted data was largely written language, but four important releases would make vast corpora of voice data available to researchers. In 2001, Apple released the Mac OS X operating system and in 2006, Microsoft released Windows Vista. Both of these operating systems contained native voice control technology (Pinola). While voice interaction was inferior to keyboard and mouse for desktops and laptops, the release of the iPhone in 2007 changed the paradigm. Google Voice Search for the iPhone was released just a year later, easing the burden of tiny keyboard buttons. In 2011, Google’s English Voice Search system recognized 230 billion words (Pinola). Today, a number of commercially available speech-recognition tools — with closely guarded intellectual property — have flooded the market including Amazon Lex, IBM Watson, Google Cloud Natural Language API, Microsoft Cognitive Services and Facebook’s DeepText.

Case Studies

Big Data at DARPA

Known for developing the precursor the internet, GPS technology and stealth planes like the F-117 Nighthawk, The U.S. Defence Advanced Research Projects Agency (DARPA) has invested heavily and openly in technology that enables the automatic transcription of speech under various conditions (Graham-Rowe). As early as 1971, DARPA sought to develop the first continuous speech transcription and granted generous contracts to institutions like IBM, Carnegie Mellon University and the University of Cambridge. The participants of the DARPA EARS (Effective, Affordable, Reusable Speech-to-Text) program required access to robust audio samples with accompanying manual translation and annotation.

With seed money from DARPA, the Linguistic Data Consortium was founded at the University of Pennsylvania in 1992. By 2004, the Consortium published more than 288 linguistic databases and established itself as a leader in linguistic big data. To support the EARS project, LDC provided audio files and careful transcriptions and pronunciation lexicons for thousands of hours of broadcast news and telephone conversations in English, Mandarin and Arabic (Strassel 1). The LDC also created "annotated corpora and guidelines to support the EARS Metadata Extraction (MDE) program" ("Past Projects").

In 1966, John Pierce — an executive at Bell Labs and the chair of the Automatic Language Processing Advisory Committee — produced back-to-back reports about machine translation and automatic speech recognition. These reports were highly critical of research underway. Pierce went so far as to write, "We are safe in asserting that speech recognition is attractive to money. The attraction is perhaps similar to the attraction of schemes for turning water into gasoline, extracting gold from the sea, curing cancer, or going to the moon. One doesn't attract thoughtlessly given dollars by means of schemes for cutting the cost of soap by 10 percent. To sell suckers, one uses deceit and offers glamor" (Liberman).

Except for a DARPA-funded attempt to master artificial intelligence — not machine translation or automatic speech recognition — that ran from 1972-1975, funding from the US government had fallen off completely. It wasn't until 1986 that researchers were able to kickstart interest in a new project, largely by making it "open-source." DARPA Program Manager Charles Wayne pitched the first of a long line of projects that included well-defined objectives, clear evaluation metrics applied by a third party, and the use of shared data sets. In the spirit of collaboration and transparency, all development training and test data was (and continues to be) published at the start of the project (Liberman).

There have been a number of DARPA human-language projects over the decades. The latest of these is the Robust Automatic Transcription of Speech program (RATS). As with projects past, the LDC supported the work of researchers at IBM, Ratheon, the Science Applications International Corporation and SRI International ("Robust Automatic Transcription of Speech") by providing 3,000 hours of "Levantine Arabic, English, Farsi, Pashto, and Urdu conversational telephone speech with automatic and manual annotation of speech segments" ("RATS Speech Activity Detection"). The goal of the RATS program is to determine whether noisy signal includes speech; identify which language is being spoken; identify whether the speaker is on a list of known individuals; and spot keywords or phrases of interest (Doermann).

Winning Big With Watson

With the exception of the current RATS program, projects at DARPA have largely been conducted before the rise of the internet and mobile technology. While the LDC provided the big data of its time (thousands of hours of recordings and careful transcriptions of those recordings), IBM developed its Watson product with expansive data from the web.

After tackling chess with a win against a grandmaster in 1997, researchers at IBM set their sights on the television trivia game *Jeopardy!.* The development of Watson and its underlying technology, called "DeepQA," represented a significant advance in machine learning (Kelly 24). By providing Watson a large corpus of unstructured information and algorithms to extract knowledge from it, the supercomputer was able to learn new material much like humans do — "by experiencing a lot of things and drawing inferences and lessons from those encounters" (Kelly 24). In preparation for the on-air competition, Watson consumed encyclopedias, dictionaries, books, news, movie scripts and more, which were stored offline for retrieval during the game. The team wrote algorithms that enabled Watson to provide answers in natural language format.

In Jan. 2011, two human contestants — former *Jeopardy!* grand champions Ken Jennings and Brad Rutter — took the stage in the auditorium of the T.J. Watson Research Laboratory in Yorktown Heights, New York. The system included 10 racks of IBM servers running Linux with the equivalent of 15 terabytes of RAM and was represented on stage by a monitor displaying a glowing blue and green avatar (Fitzgerald). Jennings and Rutter were defeated handily by the third day of competition. Watson scored $77,147 to Jennings's $24,000 and Rutter's $21,600 (Markoff). IBM donated Watson's $1 million in prize winnings to charity. "This was no mere parlor trick; the scientists who designed Watson built upon decades of research in the fields of artificial intelligence and natural-language processing and produced a series of breakthroughs. Their ingenuity made it possible for a system to excel at a game that requires both encyclopedic knowledge and lightning-quick recall" (Kelly 1).

After winning *Jeopardy!,* and with a nod to the development of MedSpeak, Watson moved on to tackle problems in healthcare. IBM is working with physicians at Memorial Sloan-Kettering Cancer Center and the Cleveland Clinic to develop a voice-driven assistant that can help diagnose diseases, identify treatment options and detect cancer. "This is more than a machine," said Dr. Larry Norton, oncologist at Memorial Sloan-Kettering. "Computer science is going to evolve rapidly and medicine will evolve with it. This is coevolution. We'll help each other."

Gigabytes at Google

If projects at DARPA represent our ability to detect and transcribe text and IBM's Watson framework represents our ability to comb mountains of data for a solution, there is currently only one company with the quantity of data required to master both. Google has become a ubiquitous global resource for questions, symptoms, needs and desires. "In 2013, Google generated 25 percent of all Internet traffic in the United States directly, and roughly 60 percent of all devices on the Internet exchanged data with Google servers on any given day" (Finn 65).

Google’s speech group was developed in 2005. In 2006, the team launched GOOG-411, a telephonic speech recognition and web search tool that helped people find directory information by speaking the city, state and name of the business they were looking for. In a little over a year, the team collected 10 million voice queries (and transcripts). They used that data as well as 2 billion Google Maps queries and 20 million entries in business databases to train their models (van Heerden 1). This immense data set, plus 250 thousand specially recorded entries, helped Google engineers prepare for their next product — voice interface for a search engine. The speech team launched Voice Search in 2008 for both iOS and Android. Rather than using screen-based buttons, users could speak their unstructured search query directly into the phone. Because the user was not making a phone call, the voice input was transmitted to Google servers via the “data channel,” dramatically improving the quality of the audio transmission (Van Heerden 2).

According to the 2001 patent for a “Voice interface for a search engine” issued to Sergey Brin and others at Google, each and every voice search helps train models for language, phonetics and acoustics. As of 2016, Google CEO Sundar Pichai said that 20 percent of all search queries are issued via voice (Sterling). That means as many as 400 billion searches each year or 1.1 billion searches every day were conducted using voice as of 2016 (Sullivan). As of May 2017, the word accuracy rate for Google Voice Search is 95 percent. That’s a 20 percent increase over Google’s accuracy rate four years earlier (Glaser).

As a natural extension of Voice Search, Google announced the launch of Google Assistant at its 2016 developer conference. The Google Assistant moves beyond simple speech transcription and begins to offer a Star Trek Computer-like experience. In contrast to Apple’s Siri, which leverages a small set of taxonomies from the web to assist users, the Google Assistant leverages the entire Google KnowledgeGraph. “KnowledgeGraph is an open ontology, drawing information from ‘controlled’ sources like Wikipedia that are primarily human-edited, but also from the unstructured data of all the web pages Google scans” (Finn 71).

Because Amazon got a two-year head start on releasing their product, and a three-year head start on selling hardware, sales of Amazon Alexa-enabled devices account for 70 percent of purchases in the category. However, critics acknowledge that Google has the technical edge. "Google Assistant has always been able to respond more fluidly to your commands. Thanks to its eponymous search engine and the data derived from it, Google is better than Amazon at answering your questions, including using the context from your previous question to help it answer your next one" (Gebhart).

The vision for Google (at large) and Google Assistant enabled products is a lofty one. Google's chairman Eric Schmidt believes, "I actually think most people don't want Google to answer their questions. They want Google to tell them what they should be doing next" (Finn 66). According to "What Algorithms Want: Imagination in the Age of Computing" author Ed Finn, "This simple comment articulates a profound shift in Google's role, one that marks their transition from the company that mastered information access in the age of the algorithm to the company that wants to build the *Star Trek* computer" (Finn 66).

What's Next

Since the historical shift from Chomskian lexicon and grammar, probabilistic modeling and machine learning have enabled impressive natural language processing. Both probabilistic modeling and machine learning require training sets — massive amounts of domain data that enable the computer system to become an expert in a given topic area. At one time, organizations like the LDC were formed to create those data sets. In today's web-connected and mobile-first environment, that data can be found — though it may come at a price.

John Kelly and Steve Hamm, authors of "Smart Machines: IBM's Watson and the Era of Cognitive Computing," lay out four Vs of big data in their book: volume, variety, velocity and veracity. While the cost of moving and transporting data has declined, the exponential increase in volume is itself a challenge. We must address this challenge and improve our ability to store data long term. Though relational databases have greatly improved our ability to store data, we must shift the paradigm as the variety of data types and sources grows. We must develop more efficient ways process data as it’s ingested, improving the velocity of actions taken in real time. So much of data captured by today’s smart devices and wearables is noise with questionable veracity; we must improve our ability to parse signal from noise (Kelly 45).

Alexa, Siri and Cortana are still in their nascent stages. To avoid embarrassing mistakes, each step forward is a cautious realization of just one feature of Jarvis, Kit, HAL or the *Star Trek* computer. But time — and an abundance of well-managed data — will ensure that one of the global tech giants will meet our science-fiction expectations and beyond.

Bibliography

“701 Translator.” IBM Archives, www-03.ibm.com/ibm/history/exhibits/701/701 _translator.html.

Bellos, David. “I, Translator.” The New York Times, 20 Mar. 2010, www.nytimes.com/2010/03/21/opinion/21bellos.html.

Doermann, David. “Robust Automatic Transcription of Speech (RATS).” Defense Advanced Research Projects Agency, www.darpa.mil/program/robust-automatic -transcription- of-speech.

Finn, Ed. *What Algorithms Want: Imagination in the Age of Computing*, MIT Press, 2017. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/newschool/detail. action?docID=4819947.

Fitzgerald, Jim. “IBM computer taking on 'Jeopardy!' champs for $1M.” Phys.org, Jan. 2011, www.phys.org/news/2011-01-ibm-jeopardy-champs-1m.html.

Gebhart, Andrew. “The Google Assistant can outsmart Alexa. So what?” CNET, 5 Oct. 2017, www.cnet.com/news/google-home-is-smarter-than-the-amazon-echo-does-it-matter/.

Glaser, April. “Google's ability to understand language is nearly equivalent to humans.” Recode, 31 May 2017, www.recode.net/2017/5/31/15720118/google-understand- language-speech-equivalent-humans-code-conference-mary-meeker.

Graham-Rowe, Duncan. “Fifty years of DARPA: Hits, misses and ones to watch.” New Scientist, 15 May 2008, www.newscientist.com/article/dn13907-fifty-years-of- darpa-hits-misses -and-ones-to-watch/.

Greenemeier, Larry. “20 Years after Deep Blue: How AI Has Advanced Since Conquering Chess.” Scientific American, 2 June 2017, www.scientificamerican.com/article/ 20-years-after-deep-blue-how-ai-has-advanced-since-conquering-chess/.

Hancox, P.J. “A brief history of Natural Language Processing.” University of Birmingham School of Computer Science, www.cs.bham.ac.uk/~pjh/sem1a5/pt1/ pt1_history.html.

Houston, Thomas. “Siri, make it so: what designers can learn from sci-Fi interfaces.” The Verge, 2 July 2013, www.theverge.com/2013/7/2/4430234/siri-make-it -so-what-designers-can-learn-from-sci-fi-interfaces.

Jones, Karen Sparck. “Natural Language Processing: A Historical Review.” University of Cambridge Department of Computer Science and Technology, October 2001, www.cl.cam.ac.uk/archive/ksj21/histdw4.pdf.

Kelly, III, John E., and Steve Hamm. *Smart Machines: IBM's Watson and the Era of Cognitive Computing*, Columbia University Press, 2013. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/newschool/detail.action?docID=1319720.

Mark, Liberman. “Lessons for Reproducible Science from DARPA’s Programs in Human Language Technology.” Stanford University, 2011, http://web.stanford.edu/~vcs/ AAAS2011/AAAS2011Liberman.pdf.

Markoff, John. “On 'Jeopardy!' Watson Win Is All but Trivial.” The New York Times, 16 Feb. 2011, www.nytimes.com/2011/02/17/science/17jeopardy-watson.html.

Manning, Christopher D. “Probabalistic Syntax.” Stanford Natural Language Processing Group, 12 Jan. 2002, www.nlp.stanford.edu/manning/papers/probsyntax.pdf.

Manning, Christopher. “Probabalistic Models in Computational Linguistics.” Stanford Natural Language Processing Group, 2000, http://nlp.stanford.edu/manning/ talks/ima2000.pdf.

“Past Projects.” Linguistic Data Consortium, www.ldc.upenn.edu/collaborations/past-projects.

Pieraccini, Roberto. “From AUDREY to Siri: Is speech recognition a solved problem?” International Computer Science Institute at Berkeley, http://www.icsi.berkeley .edu/pubs/speech/audreytosiri12.pdf.

Pinola, Melanie. “Speech Recognition Through the Decades: How We Ended Up With Siri.” PCWorld, PCWorld, 2 Nov. 2011, www.pcworld.com/article/243060/spe ech_recognition_through_the_decades_how_we_ended_up_with_siri.html.

“Pioneering Speech Recognition.” IBM100 - Pioneering Speech Recognition, www-03.ibm.com/ibm/history/ibm100/us/en/icons/speechreco/.

“RATS Speech Activity Detection.” Linguistic Data Consortium, 2015, www.catalog.ldc.upenn.edu/ldc2015s02.

Rao, Gauri, et al. “Natural Language Query Processing on Dynamic Databases Using Semantic Grammar.” International Journal on Computer Science and Engineering, vol. 2, no. 2, 2010, pp. 219–223, www.enggjournals.com/ijcse/doc/IJCSE10-02- 02-20.pdf.

“Robust Automatic Transcription of Speech (RATS).” Electronic Privacy Information Center, epic.org/foia/darpa/rats/.

Schank, Roger C. “Language and Memory.” Cognitive Science, vol. 1, no. 4, 1980, pp. 243–284, www.web.stanford.edu/class/linguist289/schank80.pdf.

Strassel, Stephanie. “Linguistic Resources for Effective, Affordable, Reusable Speech-to-Text.” Language Resources Evaluation Conference, 2004, Linguistic Resources for Effective, Affordable, Reusable Speech-to-Text .

Sterling, Greg. “Google says 20 percent of mobile queries are voice searches.” Search Engine Land, 24 May 2016, www.searchengineland.com/google-reveals-20- percent-queries-voice-queries-249917.

Sullivan, Danny. “Google now handles at least 2 trillion searches per year.” Search Engine Land, 14 Mar. 2017, www.searchengineland.com/google-now-handles-2-999- trillion-searches-per-year-250247.

Terras, Melissa M., et al. *Defining Digital Humanities: A Reader.* Routledge, 2016. https://books.google.com/books?id=xAYpDAAAQBAJ&pg.

van Heerden, Charl, et al. “Language Modeling for What-with-Where on GOOG-411.” International Speech Communication Association, 2009, www.isca-speech.org /archive/archive_papers/interspeech_2009/papers/i09_0991.pdf.

Wichmann, Anne. “Teaching and Language Corpora.” Routledge, 11 June 2014, https://books.google.com/books?id=eG7JAwAAQBAJ&pg.

Winograd, Terry. “SHRDLU.” Stanford HCI Group, www.hci.stanford.edu/winograd/ shrdlu/.