Chatbot Based Content Discovery: Faulknerbot in the Archive

Aaron Mauro
mauro@psu.edu
Pennsylvania State Digital Humanities Lab, United States

Introduction

In March of 2016, the failure of Microsoft's prototype chatbot, Tay, was not just a technological failure. It was a disciplinary failure. It was a failure of an industry leader to adopt a critical perspective when building systems in a complex cultural and social environment. Tay, which stands for "thinking about you," was the name given to an artificial intelligence chatbot for Twitter that was quickly corrupted by users and began spewing racist, sexist, and homophobic slurs. Pundits quickly leapt to conclusions about the political beliefs of internet users, but these same pundits failed to understand that this hacking of Tay was in fact a critique of chatbots in the real world. Users of Twitter were exposing a fundamental error made by the Microsoft development team. Because the system learned direct-ly from user input without editorial control or content awareness, Tay was quickly trained to repeat slurs by users eager to embarrass Microsoft.

This moment in technological development makes for an interesting anecdote, but it also represents the moment that chatbots entered the public consciousness and became nothing less than the future direction of a unified interface for the whole of the web. Of course, chatbots captured imaginations in the 90s as well. Systems like Cleverbot, Jabberwacky, and Splotchy were fascinating to play with, but they had no real application. Today, text based AI has been identified as the the successor to keyword search. No longer will we plug in keywords into Google, comb through lists of text, and depend on search engine optimization (SEO) to deliver the best content. Search will be around for a long time, but in the near future much more content will be delivered through text based messenger services and voice controlled systems. We've seen the early stages of this change in products like Amazon's Alexa, Apple's Siri, Google Now, and Microsoft's Cortana. There are now bots embedded within common platforms like Slack, Skype, and Facebook Messager. We are now approaching a world that Apple envisioned in 1987 with a mockup system called the "Knowledge Navigator" that sought to give users an interactive and intelligent tool to access, synthesize, present, and share information seamlessly.

Humanities in the Loop

We are likely decades away from a true "knowledge navigator," but the second generation of these chatbots are now in development. The company that developed Siri for Apple is now in the final stages of development on a system called Viv (Matney). Viv is the first viable company to produce a unified interface for text and speech based AI assistants. Facebook is testing project M within its messenger app to allow users to issue commands, access services, and make purchases through text input (Hempel). The remarka-ble thing about M is that Facebook has built a system with "humans in the loop." This means that when a service is accessed, perhaps by purchasing movie tick-ets, a human will fine tune the AI generated results for each transaction. There is currently an understanding within the machine learning community that human assisted training of these systems produces more accurate results but will also train more robust systems going forward (Biewald, Bridgwater). The current need for human in the loop systems means that we are at a crucial moment for humanists to lend their experience and critical abilities to the development and training of AI systems. In the field of machine learning, training a system to answer humanities based problems will show how these systems succeed or fail, but they will also demonstrate the value of the humanities in a digital world. If the purpose of the humanities is to better understand what it is to be human, training AI to answer philosophical, historical, or cultural questions will help us understand our experiences as we become more accustomed to intelligent systems in our lives. Grappling with AI, whether it is in a mundane consumer exchange or in matters of grave ethical importance, is rapidly becoming a practical problem in our lives.

With humanists in the loop, we will better understand the social and cultural contexts in which these systems appear and avoid the regrettable failure of systems like Tay in the future. We are currently on the cusp of a revolution in the applicability of natural language understanding, artificial intelligence, and conversation based interfaces design. These technologies will have ranging consequences socially, culturally, and economically in the coming decade, but these technologies are also deeply connected to the social and cultural contexts in which they appear. My goal is to train

machines to be humanists. It is the literary crit-ic's ability to close read complex philosophical, histori-cal, and artistic meaning that these systems lack. It is the ability of the historian to contextualize political and technological change within the breadth of human progress. It is the dramatists ability to understand performance and dialogue that will animate our conversations with computers. The digital humanities are well situated to make the most of NLP techniques and find culturally significant training sets.

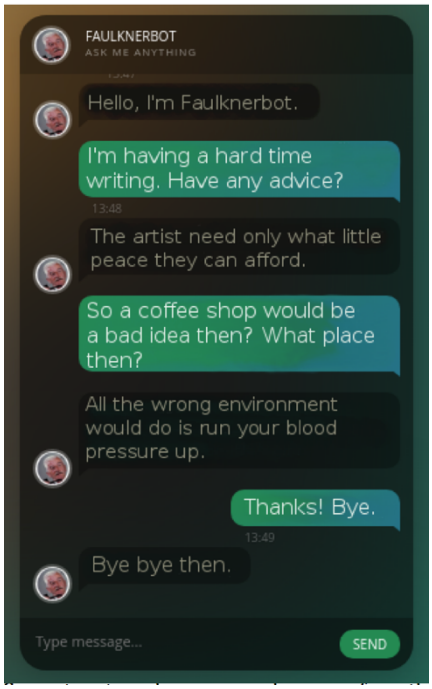

Figure 1: Faulknerbot interface with basic query and response

Method: Conversational Data Retrieval

Biographical and archival material has been used to train a system to allow a conversation with the famed American author William Faulkner. I will pre-sent a system trained with nearly all the interviews that Faulkner has given. Author interviews are an ex-cellent training set because the questions asked by the interview anticipate user interests and model a con-versational style of response. The interviews collected during Faulkner's visit to the University of Virginia were instrumental in building this tool. The applica-tions for such a system are numerous. A conversation with Faulkner might benefit a creative writing student in the midst of writer's block. A chatbot offers a more inviting interface for a general public. Most important-ly, Faulknerbot will represent a novel form of content discovery for student researchers. Once a user has developed a chat history worth exploring, Faulknerbot's responses link to original archival materials for research purposes.

Current systems have come a long way from the toy-like chatbots that populated the web in the late 90s. After a pre-processing stage using word2vec, which vectorizes the bag of words, this model uses Tensor Flow to generate two complementary neural networks that encodes and decodes inputs and responses. This model has only recently been made accessible to non-computer science researchers recently by Google open sourcing Tensor Flow. is not based on the retrieval based model using a rule based expressions, with a heuristic to determine intent and draw from a predefined response. This is not a simplistic tree model based on nested "if/then" statements. Instead this uses a generative model. This generative model uses sequence to sequence learning with neural networks (https://arxiv.org/abs/1409.3215). This model links words statistically to determine "flows" of meaning through a word vector. Geoff Hinton calls this a "thought vector." In other words, this is an end-to-end model that remains open. Rather than a retrieval method, which limits the scope of the conversation, this system dynamically learns and allows for a retention of what has been said. The generative model allows for this context based discussion without resorting to an enormous conversation log. In Tensor Flow, this operates on a Long Short Term Memory (LSTM) network. As I've said, the sequence to sequence model is based on two neural nets. One is an encoder, which encodes input data from the user. The decoder model determines the reply by generating the output, which need not echo the size of the vector. This thought vector generalizes input and links to a target response. This is not a "feed forward" neural net. It is a recurrent neural net that continually retrains on the training data, which is often the marker of a true "deep learning" system. This model makes no assumption about purpose or predetermined output. It simply reinforces relationships between thought vectors over time. There is a deeply emotional resonance that is carried through conversation. The blurring of lines between social media, search, and messaging will result in a seamless and unified interface for digital technology. Driven by the mobile space's demand for streamlined UI design, we will become more reliant on assistive technologies that can anticipate, learn, and adapt to user input.

Conclusion

It is important for the humanities to anticipate this new cultural space. When the Google autocomplete

system was introduced to search, there were many cultural commentators decrying the loss of independ-ent thought and the potential for entrenching damag-ing stereotypes (Postcolonial). The loss of critical awareness and even just the ability to spell. Technolo-gy that offends our sense of what it is to be essentially human is usually the next important media type. Chat-ting with machines tends to cross such lines. There are practical uses for remedial education and composition studies. A functioning Teaching Assistant Bot capable of answering questions about deadlines, assignments, and course policy would be welcome by most educators. Indeed, an AI TA has been developed recently, but it is unclear if this system can be trained on any course material or was custom built for this class (Maderer). Generalizing these systems is a difficult task, to be sure. The newly open sourced Tensor Flow machine learning library can answer questions derived from a training set of just over a million words. When we consider the limits of machine learning in intelligent assistants, scholarly communication through chat interfaces is certainly the next logical step. However these systems require humans in the loop. They requirethoughtful and critical reflection. They require an attention to depth and nuanced meaning. They require a humanist in the loop.

Bibliography

Alphabet. (2011) Google Code Archive. Web. <https://code.google.com/archive/p/word2vec/>.

Apple (2016) "Knowledge Navigator." 28 October. <https://www.youtube.com/watch?v=HGYFEI6uLy0>.

Bridgwater, A. (2016) "Machine Learning Needs A Human-In-The-Loop." 7 March. <http://www.forbes.com/sites/adrianbridgwater/2016/03/07/machine-learning-needs-a-human-in-the-loop/#2175c4ba6590>.

Biewald, L. "Why human-in-the-loop computing is the future of machine learning." 13 Nov. 2015.<http://www.computerworld.com/article/3004013/robotics/why-human-in-the-loop-computing-is-the-future-of-machine-learning.html>.

Carpenter, R. Cleverbot. 28 Oct. 2016. <http://www.cleverbot.com/>.

Carpenter, R. Jabberwacky. 28 Oct. 2016. <http://www.jabberwacky.com/>.

Railton, S. et al. (n.d.)Faulkner at Virginia. <http://faulkner.lib.virginia.edu/>.

Hempel, J. "Facebook Launches M, Its Bold Answer to Siri and Cortana." 26 Aug. 2015. <https://www.wired.com/2015/08/facebook-launches-m-new-kind-virtual-assistant/>.

Hunt, E. (2016). "Tay, Microsoft's AI chatbot, gets a crash course in racism from Twitter." 24 March .Web. <https://www.theguardian.com/technology/2016/mar/24/tay-microsofts-ai-chatbot-gets-a-crash-course-in-racism-from-twitter>.

Maderer, J. (2016) "Artificial Intelligence Course Creates AI Teaching Assistant." 9 May 2016. <http://www.news.gatech.edu/2016/05/09/artificial-intelligence-course-creates-ai-teaching-assistant>.

Matney, L. (2016). "Siri-creator shows off first public demo of Viv, 'the intelligent interface for everything'" TechCrunch. 9 May.<https://techcrunch.com/2016/05/09/siri-creator-shows-off-first-public-demo-of-viv-the-intelligent-interface-for-everything/>.

Microsoft. (2016) "TayTweets." Twitter. 28 October. <https://twitter.com/tayandyou>.

Postcolonial Digital Humanities. (2013) "Google's Auto-com-pletion: Algorithms, Stereotypes and Accountability." 19 November. <http://dhpoco.org/blog/2013/11/19/googles-auto-completion-algorithms-stereotypes-and-accountability/>.

Splotchy (2016). Algebra.com 28 Oct. <https://www.algebra.com/cgi-bin/chat.mpl>.