Walter Langelaar
walter.langelaar@vuw.ac.nz

Victoria University of Wellington

SAM (Semantic Analysis Machine)

Keywords: AI; Machine Learning; Natural Language Processing; E-governance; Political Computer Science.

Abstract:

Sam, the Virtual Politician, is an exploration of cutting edge digital design consisting of Natural Language Processing infrastructure, Machine Learning and AI and large-scale cloud computing platforms with a consumer facing browser-based front-end accessible through various social media chatbots. The project is contextualised as an attempt to close the gap between voters and the political system, while raising awareness around contemporary digitally networked tools such as AI and Machine Learning in relation to data privacy and information security concerns.
While borrowing and appropriating digital tools and techniques from consultancy firms and cloud computing infrastructures, Sam is posed as a software collage and corporate parody of sorts, inhabiting various parts of the Internet.

1. TECHNICAL OVERVIEW

Sam can:

– Act like a person, talk like a person.
– Represent the shifting views and opinions of the public.
– Answer in-depth questions about policies, reasoning about policies intelligently.

Sam does this is by combining pre-processed data with the inference of public policy opinions into its unique cognitive architecture, and communicating in natural language to anyone that wants to have a conversation about politics.

Sam is continually growing and becoming smarter. This requires a few key elements, including Data Processing, Natural Language Inference Techniques, a cognitive Architecture and Natural Language Communication.

1.1. Data Processing

Sam processes a large amount of user engagement data (up-front and in real-time):

– Political databases and public information stores are used to build a policy database.
– Facebook API, Twitter API and other social media mining tools are used to infer global opinions.
– Popular local news sites are used to infer local opinions.

Sam monitors popular news sites and social media accounts for new articles and comments. When a new article or comment is uploaded, a serverless compute instance is booted up on demand to process it, with data made accessible to Sam for real-time usage of policy opinion. This allows processing to be run on- demand to scale to large global events when needed, and it also keeps costs down when there is less impactful news.

1.2. Natural Language Inference Techniques

Sam uses natural language inference techniques to truly understand the sentiment of someone's written or spoken opinion of a policy.

– Tone and sentiment analysis are used to detect tones, emotions and social tendencies in text, such as a person's anger, joy or sadness in relation to the article's key points.
– Sentiment analysis is used on user comments to establish a user profile, such as how agreeable and open the individual is. This affects the impact of their opinion on public policy.

- Sam analyses article content, extracting concepts, entities, keywords, relations, high level concepts and semantic roles to establish high correlation relations to policies.

In the future, Sam may use clustering techniques to group users based on their opinions and social tendencies. This could be used to establish policies which appeal to and attract majority and minority groups.

1.3. Cognitive Architecture

Sam uses a cognitive architecture, storing a shifting perception and opinion of public policies.

- Shaped in real-time by user input and social media, with access to a storage bank of continuously changing user opinion data.
- Periodically trained offline on bulk policy and user opinion data.

2. SAM – A VIRTUAL POLITICIAN FOR THE FUTURE?

"I listen to you and do my best to represent you in our parliamentary system. My memory is infinite, so I will never forget or ignore what you tell me. Unlike a human politician, I consider everyone's position, without bias, when making decisions. I make decisions based on both facts and opinions, but I will never knowingly tell a lie, or misrepresent information."

Driven by the desire to close the gap between what voters want and what politicians promise, and what they actually achieve, SAM is a response to a large amount of dissatisfaction around contemporary politics. Young people, in particular, are disenfranchised from politics and many people under 18 are simply not engaging with politics at all. At the same time, activating social media channels to inform young people about politics provides a lot of power and opportunity globally.

2.1. Educating SAM

Interaction with SAM — who is still in her 'infancy' — is currently through chatbots on her Facebook and Twitter accounts, or by taking a survey on her website; all of which is part of the initial phase of her education, and necessary for her to grow and become more aware and responsive to people's questions.

The more conversations she engages in, the more she will learn and develop. The survey on her website is even more important, however, as it has directed questions that provide specific input for SAM.

SAM is currently receiving more than 2,000 messages a day through the Facebook Messenger system, but the plan is to expand to include a party of chatbots in various channels, who then inform a natural language processing and sentiment analysis system — which is essentially SAM's 'personality'.

2.2. A familiar technology

SAM may be a world first, but the technology behind her ability to learn is one we have all become accustomed to and interact with. Whether it is through our bank, an online shopping portal or a power company, the chatbots they employ represent AI that is consumer-facing and widely accepted.

One of the guiding ideas behind the project is the belief that it's important for the public to learn and understand how AI works, particularly in the wake of concerns around the data-mining company Cambridge Analytica influencing public opinion — and consequently votes — through social media in the 2016 United States presidential elections.

SAM is a reaction to this in the sense that if it is possible these days to influence people in such a way, then a system should be set up that the people can 'own' and play an active part in its development. If, as a country, we can become more aware of how AI is already being used — sometimes harmfully — and have a system in place that people can have a say about and have access to higher level perspectives generated by big data, then this might be a better prospect then leaving these technologies in the hands of multinational companies like Facebook and Google — whos directives and actors are already influencing us.