HHS: Working Smarter, Not Harder with AI and NLP

HHS Office of the Chief Technology Officer - August 2019
Alex Curtis - M.S. in Transportation Planning and Engineering at New York University

imcurtis@gmail.com AND Alexander.Curtis@hhs.gov

Why I am here (HHS) this summer

As part of the Civic Digital Fellowship, I have been selected from a large pool of applicants to apply my skills and experience in **data science**, **communications**, and **civic engagement** to increase public awareness and involvement with the government, with Lyme disease and the **Tick-Borne Disease Working Group** as a use case.

The Problem

HHS receives thousands of public comments and unstructured text from rule making and its hundreds of FACA working groups, like the Tick-Borne Disease Working Group (TBDWG). This results in 1000s of submissions that HHS manually reviews, which cost time, resources, and taxpayer dollars.

Proposed Solution

Emerging technology solutions exist. Artificial intelligence (AI) and natural language processing (NLP), for example, can mine millions of public comments and efficiently distill the textual data to identify key themes/topics as outputs.

Background

Artificial intelligence, or AI, is intelligence demonstrated by **machines** as opposed to the **natural intelligence** of humans.

- Describes machines that mimic certain human cognitive functions (e.g. problem solving)

NLP is a component of AI

- Focus: ability to understand human language as it is spoken
- Two general analytical techniques:
  - Semantic: understand meaning and structure of sentences
  - Syntactic: assess meaning from a language based on grammatical rules

My methodology

Using specific techniques to achieve the goal: automating the review process for public comments

Analysis split into four parts:

1. Structuring and preprocessing text data for later analysis
2. Topic modeling: identifying a certain number of distinct themes from the comments
3. Sentiment analysis: identifying how each respondent feels given their written comments
4. Explore how sentiment changes over time - before and after the publication of TBDWG 2018 Report to Congress on November 15, 2018

Figures (workflow)

Figure 1. Topic Portion. Diagram that explains the flow of information from text data to resulting topic

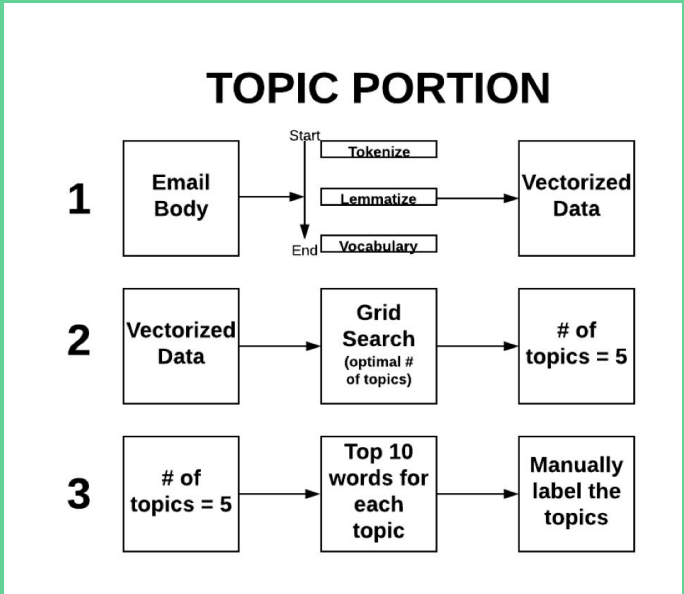
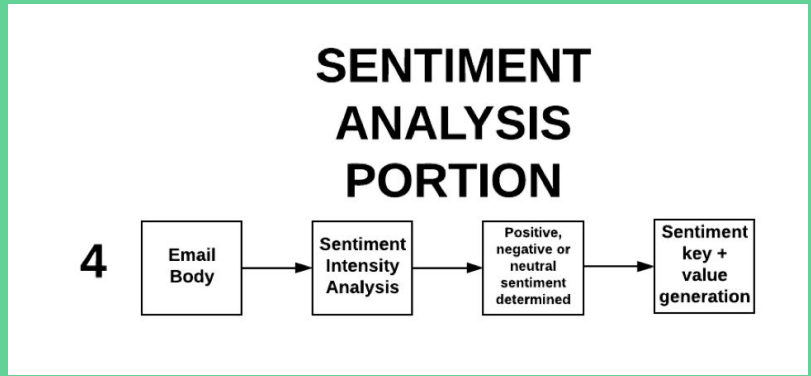


Figure 2. Sentiment Analysis Portion. Diagram that explains flow of information, from email body to sentiment (positive, negative or neutral) generation



Outputs

Table. 1: Sentiment Analysis Results						
Theme	**Positive Before 11/15/18**	**Neutral Before 11/15/18**	**Negative Before 11/15/18**	**Positive After 11/15/18**	**Neutral After 11/15/18**	**Negative After 11/15/18**
Patient experience	32%	21%	47%	13%	17%	70%
Public engagement	56%	26%	18%	56%	24%	20%
Physical manifestation of disease	52%	36%	12%	0%	100%	0%
Doctor-patient relationship	49%	21%	30%	33.30%	33.30%	33.30%

Recommendations

1. Submit a report to HHS CTO with recommendations on the long-term viability of this type of NLP project, including: how to structure the input data and produce output data through automation, as well as finding and applying different methods of sentiment analysis

2. Conduct an assessment on how this NLP technology may be applied to create HHS efficiencies and cost savings. Assess interagency need and technology potential within federal government, as well as the current use and best practices from industry.

3. Submit 10x proposal (done) for GSA support to further scope and potentially scale NLP for public comment submissions, which will save HHS time, resources, and taxpayer dollars.

10x Funding Opportunity

- Incremental investment fund from within U.S. federal government
- Building technologies for internal projects with an aim to improve how citizens interact with government

Idea: NLP: Nation Listens to People

Pitch: HHS has observed how it takes enormous amounts of staff time, resources, and taxpayer dollars to manually analyze written public comments submitted through rule making, FACA committees, FOIA requests for information, and other channels. Given our experience with the CMS-funded MITRE Public Comment Tool (IMPACT) for automated workflow and the results of an HHS Office of the CTO 2019 pilot on NLP, we believe that NLP and AI can semi-automate, streamline, and expedite the public comment process. To scale across HHS and government, these emerging technologies necessitate legal and policy guidance – perhaps also a playbook with “bright spot” examples and open-source tools – on how to responsibly leverage NLP and AI for written public comments to improve federal responsiveness to citizens and save taxpayer dollars.

Conclusion

This HHS CTO pilot is viable, and yielded efficiencies that we recommend be further developed and scoped out for all written public comments to HHS, including FACA working groups and rulemaking. Noted is the increase in negative sentiment after the publication of the TBDWG report to Congress. Scaling this pilot will require resources, and one potential source is GSA’s 10x funding opportunity.

Acknowledgements

Kristen Honey - Innovator in Residence with HHS CTO

Ed Simcox - HHS CTO

Margeaux Akazawa - Internal Innovation Programs Lead, Office of the CTO

James Daniel - Public Health Coordinator / Office of the National Coordinator for Health IT at HHS

Thank You!


Uppsala universitet 1852–1916
Det gamla universitetet. Festskrift till Anders Hallberg
FOLKERA CARICER
Studenten, Staden och Samingen
UPPSALA UNIVERSITE