



ENVIRONMENTAL POLICY
INNOVATION
CENTER



BlueConduit

Analyzing Public Comments on the Proposed
Changes to the Lead and Copper Rule using Natural
Language Processing



Suggested citation:

Jared Webb, Cooper Stansbury, Sridhar Vedachalam, Ian Robinson, William Nicholas, Jacob Abernethy, Eric Schwartz, and Timothy Male. 2020. "Analyzing Public Comments on the Proposed Changes to the Lead and Copper Rule using Natural Language Processing." Environmental Policy Innovation Center, Washington D.C. and Blue Conduit, Ann Arbor, MI.

Report is open-access under [CC BY-NC 4.0](#).

Authors:

Jared Webb¹

Cooper Stansbury^{1,2}

Sridhar Vedachalam, PhD³

Ian Robinson¹

William Nicholas^{3,4}

Jacob Abernethy, PhD^{1,5}

Eric Schwartz, PhD^{1,2}

Timothy Male, PhD³

¹BlueConduit, ²University of Michigan, ³Environmental Policy Innovation Center, ⁴Indiana University, and ⁵Georgia Tech.

For more information, email sri@policyinnovation.org or call 614.364.3414.

Report design by Rachel Garofolo

The [Environmental Policy Innovation Center](#) builds policies that deliver spectacular improvement in the speed and scale of environmental progress. A nonprofit start-up, EPIC is committed to finding and highlighting the best approaches to scaling up results quickly. EPIC focuses on clean water, endangered species, environmental markets and the use of data and technology in producing conservation outcomes. Our work in water focuses on innovative financing, outcomes-based stream and wetland restoration, water quality partnerships, utility consolidation, and the role of data technology in improving consumer trust.

[BlueConduit](#) leverages machine learning to ensure that residents of every community regardless of means have access to clean water. Growing out of research at the University of Michigan starting in 2016, BlueConduit has used predictive models to help cities identify which homes have lead service lines, providing a home-by-home inventory to guide their removal in a cost-effective way. BlueConduit is working with a growing list of utilities and municipalities throughout the United States.

Support for this report was provided by the Robert Wood Johnson Foundation.



HIGHLIGHTS

- Machine analysis of comments and attachments submitted to EPA found detectable differences between different agency types, but that these differences do not always correspond to differences of opinion.
- Sentiment analysis found significant differences in the linguistic content of comments and attachments; the comment text appealed to emotion, whereas attachments appealed to logic/reason.
- The documents (comments and attachments) were largely neutral in tone, with only a slightly positive sentiment.
- Submissions differed in how they presented their agenda but are reasonably consistent about what that agenda is.
- Natural Language Processing is an exciting tool to process a large set of text, but currently available techniques are not sufficient to parse language and tone to provide decisive conclusions about the authors' intent and judgment.

SUMMARY

Federal and other government agencies often take public comment on major proposed regulatory actions and policies, and this public input is considered a fundamental part of the functioning of a democratic government. In a review of major federal regulatory actions last decade, the Government Accountability Office (GAO) found that agencies only allowed public comment on about two-thirds of major rules and, of those, only responded to public comment about two-thirds of the time.¹ Electronic means of public comment have greatly increased the number of public comments on a small subset of rules, but had little difference on the median number of comments received.^{2,3} Politically controversial rulemakings are among the most likely to attract significant comment, but there is not much evidence that the volume of comments is correlated with any likelihood an agency changes its proposed action before finalizing it.⁴

Natural language processing, which refers to performing data science on textual data, can help agencies make sense of and act upon public comment as well as help advocates and the legal system themselves understand public comment for the purpose of evaluating the adequacy of agency response. It relies on techniques like topic analysis, in which a large collection of text is used to find a specified number of repeated themes and the keywords that define those themes, and sentiment analysis, whereby one can estimate the speaker's feeling on a 'positive/negative' spectrum in a given passage of text, to disaggregate large volume of text and find patterns that are too laborious, if not outright impossible, to detect by humans.

We used natural language processing techniques to evaluate methods for sentiment and topic analysis, and to identify areas for future policy research. We examined public comments on the proposed revisions to the Environmental Protection Agency's (EPA) rule on lead and copper regulation in drinking water proposed in 2019. The rule attracted 80,000 submitted comments of which approximately 8500 were not identical form letters. This action was extremely important to the public since the Flint (Michigan), Washington (DC), Newark (New Jersey) and other water utilities' drinking water contaminant crises have created widespread public concern about the safety of drinking water. Both the Trump and Obama Administrations had postponed releasing the draft regulations multiple times.

Automated sentiment and topic analysis found no subject areas that would not be detectable to a human reader and this analysis offered little to suggest that automated analysis would be useful to agencies in determining what the public is interested in. However, we found the analysis offered insights that challenged the biases with which human readers probably often review comments. For example, machine-distinguishable patterns between state agency comments and water utility comments highlighted differences that we believe EPA should look for in its public comment analyses. Children's health was a more common theme than we anticipated. Initiatives around lead service line replacement appeared to be more positively received than other areas in which the proposed regulations attracted interest. The very low mention of cost concerns from the public suggest that the federal government should not exaggerate cost considerations in its final rule, compared to the draft.

We believe that these kinds of insights suggest that there already is - or soon will be - value in the government building automated sentiment and topic analysis into the public comment response procedures of federal agencies when they receive high volumes of unique and detailed public comments (say, more than the 95th percentile in the previous 5-year period).⁵ Such analyses are likely to pick up patterns in public comment that could help make agencies' response to comment more accurate and less subject to bias. While we did not find the technology to be easy to apply, analyses like those of FiscalNote show that there is already value in machine learning-based analysis for agencies like the GAO, Congressional Research Service, or White House Office of Management and Budget which look at patterns across the federal government and over time. We expect that rapid development of machine learning technologies will soon make them invaluable for individual rule review and as a tool for courts to determine whether agencies have adequately addressed major elements of public comment and interpreted those comments accurately.

¹ General Accountability Office. 2012. Federal Rulemaking: Agencies Could Take Additional Steps to Respond to Public Comments. GAO-13-21. Accessed at <https://www.gao.gov/assets/660/651052.pdf>

² Balla, S. J., & Daniels, B. M. 2007. Information technology and public commenting on agency regulations. *Regulation & Governance*, 1(1), 46-67. DOI: 10.1111/j.1748-5991.2007.00005.x

³ Dermody, A. 2020. How Regulatory Comments Have Changed from Obama to Trump. FiscalNote. Accessed at: <https://fiscalnote.com/blog/how-regulatory-comments-have-changed-from-obama-to-trump>

⁴ Shapiro, S. 2008. Does the amount of participation matter? Public comments, agency responses and the time to finalize a regulation. *Policy Sciences*, 41(1), 33-49.

⁵ The 95th percentile figure for comments received by EPA in the last three years was 110. See Dermody (2020) referenced in the report.

INTRODUCTION

Under the Administrative Procedure Act of 1946 (P.L. 79-404), each agency “shall give interested persons an opportunity to participate in the rule making through submission of written data, views, or arguments with or without opportunity for oral presentation. After consideration of the relevant matter presented, the agency shall incorporate in the rules.” While this was previously administered through hearings and other forums, a wide expansion of written notice-and-comment rulemaking began in the 1970s in which agencies posted notifications of upcoming rules and invited public comment, and has accelerated with the wider adoption of electronic communications. Since its launch in 2003, the website www.regulations.gov has enabled the public to submit comments on proposed federal regulations via the internet, and more recently has facilitated access and analysis via its developer API. It supports more than 90% of federal rulemaking, accounting for hundreds of thousands of comments on thousands of rules each year.⁹

FEDERAL REGULATIONS & PUBLIC COMMENTS

In a review of major federal regulatory actions last decade, the Government Accountability Office (GAO) found that agencies only allowed public comment on about two-thirds of major rules and, of those, only responded to public comment about two-thirds of the time (**Figure 1**).¹⁰ When public participation increases, the work of federal agencies to process these comments is more difficult to manage. The proposed changes to the Federal Communications Commission’s (FCC) policy on net neutrality garnered more than 23 million responses¹¹, and the U.S. Environmental Protection Agency’s (EPA) controversial Waters of the United States (WOTUS) recodification received nearly 800,000.¹² These two regulations are exceptions rather than the rule. An analysis of public comments submitted on regulations.gov regarding EPA regulations in the past six years (last three years of the Obama administration and the first three years of the Trump administration) suggests that a typical EPA regulation attracts merely three comments (**Figure 2**).¹³

A TYPICAL EPA REGULATION ATTRACTS MERELY THREE COMMENTS

⁶ Public comments make a difference. Regulations.gov.

Accessed at: https://www.regulations.gov/docs/FactSheet_Public_Comments_Make_a_Difference.pdf

⁷ Livermore, M. A., Eidelman, V., & Grom, B. 2017. Computationally assisted regulatory participation. *Notre Dame L. Rev.*, 93, 977.

⁸ Open Government Initiative: Regulations.gov Exchange.

Accessed at: <https://obamawhitehouse.archives.gov/open/innovations/Regulations-gov-Exchange>

⁹ Ibid.

¹⁰ General Accountability Office. 2012. Federal Rulemaking: Agencies Could Take Additional Steps to Respond to Public Comments. GAO-13-21. Accessed at: <https://www.gao.gov/assets/660/651052.pdf>

¹¹ https://www.fcc.gov/ecfs/search/filings?proceedings_name=17-108

¹² <https://www.regulations.gov/docket?D=EPA-HQ-OW-2017-0203>

¹³ Dermody, A. 2020. How Regulatory Comments Have Changed from Obama to Trump. FiscalNote.

Accessed at: <https://fiscalnote.com/blog/how-regulatory-comments-have-changed-from-obama-to-trump>

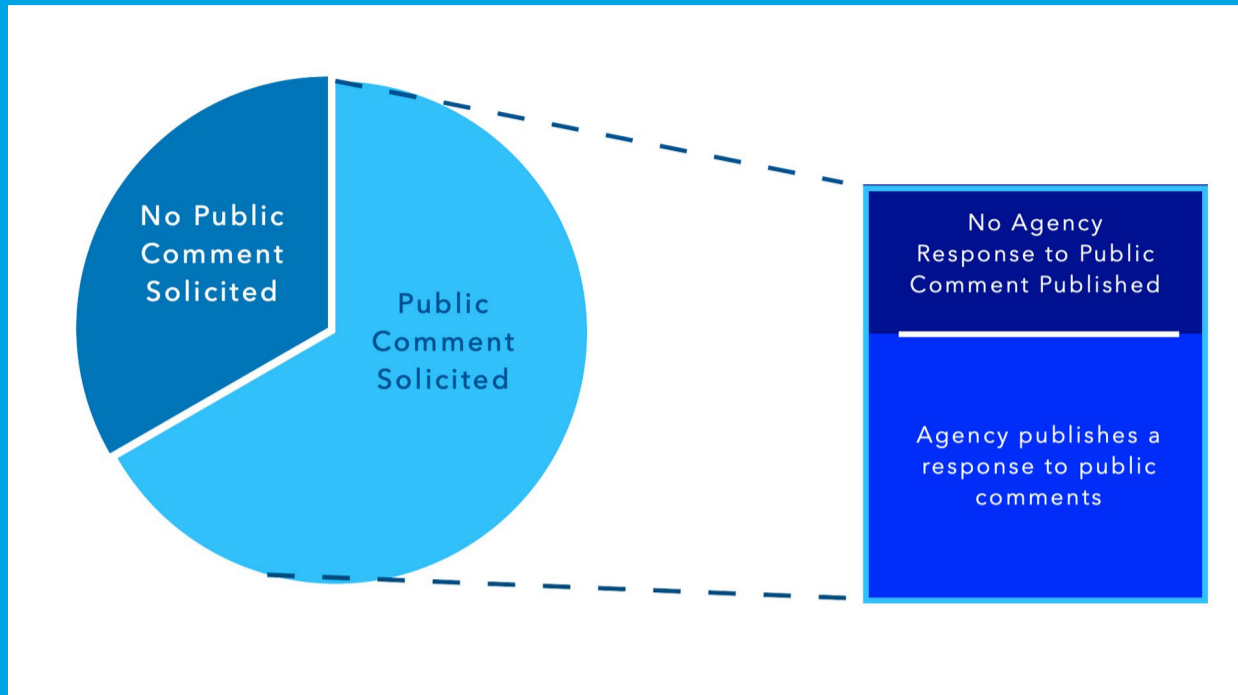


FIGURE 1

Findings from a GAO report suggest that government agencies are failing to invite and respond to public comments in a majority of major regulations.

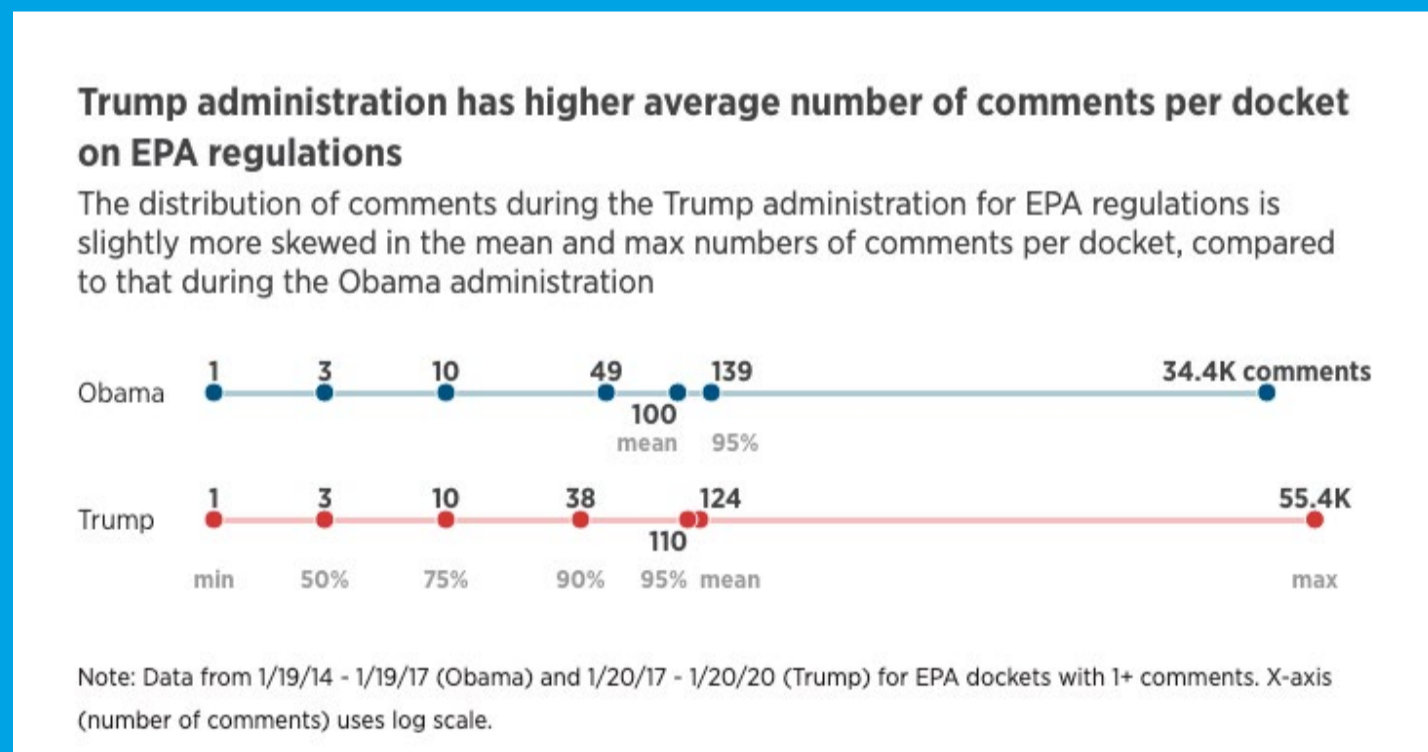


FIGURE 2

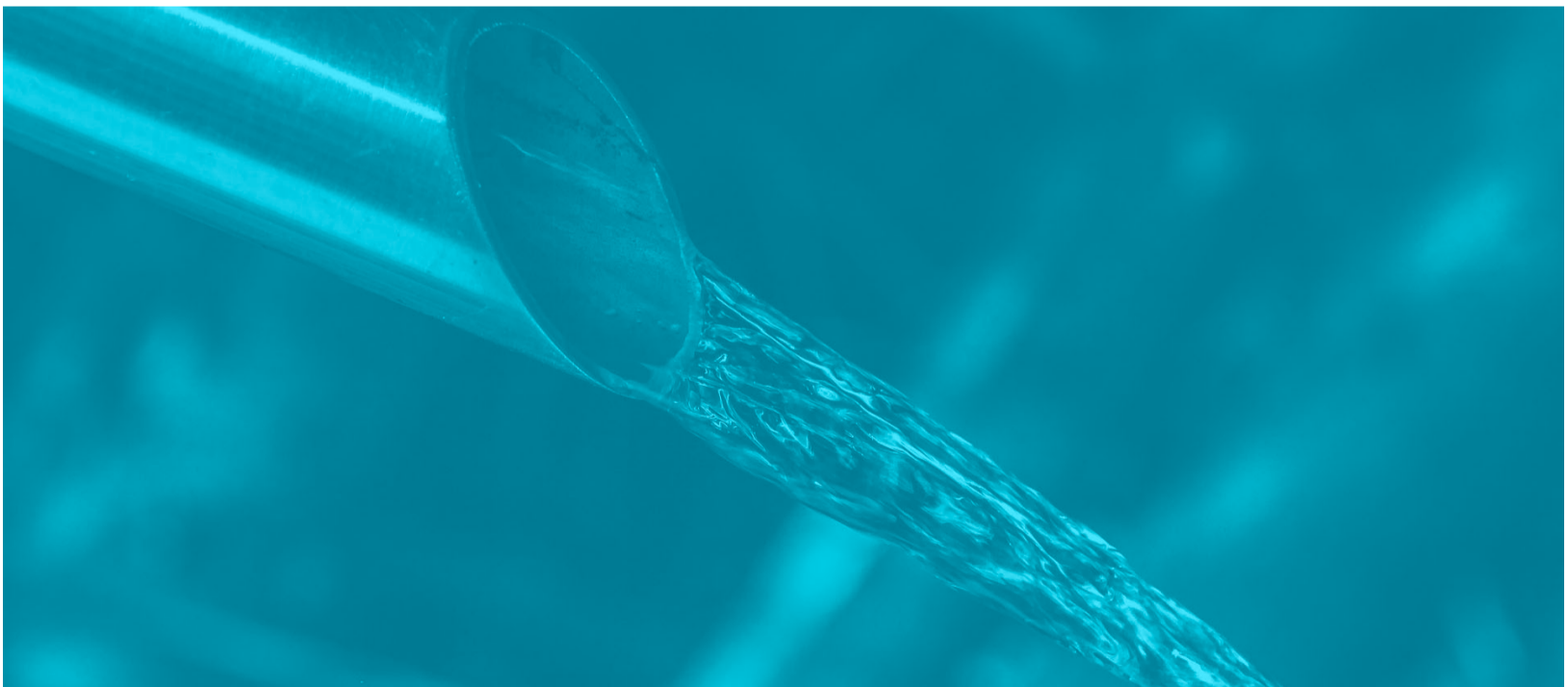
A typical EPA regulation attracts only three public comments. The average number of comments on any given regulation is closer to 100 due to outlier regulations such as WOTUS (the original rule as well as its repeal). This chart excludes the proposed Lead and Copper Rule Revisions. **Source: FiscalNote**

Regulations attracting high number of comments are likely to be either technically complex or politically controversial - or both. In theory, agencies should have a desire and need to deal with that volume of comments with accuracy and sensitivity, but still be able to manage them fast enough to meet agencies' internal and external deadlines. These large volumes of comments present federal agencies with the need for tools to a) mechanically or electronically process comments, b) exclude invalid comments, and c) process comment materials for significant input or suggestions and categorize and sort that input for response.¹⁴

EPA'S LEAD AND COPPER RULE

Lead in drinking water is regulated under a few key federal laws and regulations that include the Safe Drinking Water Act (SDWA), the Lead and Copper Rule (LCR) first promulgated in 1991, and various state and local laws focused on lead monitoring and reporting requirements at schools and child care centers and their subsequent revisions. The 1991 federal rule went through minor amendments in 2000 but lead standards have remained unchanged for years. On October 10, 2019, the EPA announced the first major revision of the federal Lead and Copper Rule (LCR) since 1991. This action was much-awaited since the Flint water crisis in 2015, and both the Trump and Obama Administrations had postponed releasing the draft regulations multiple times.

The current Lead and Copper Rule requires public water systems to monitor for lead in drinking water and for large water systems to provide treatment for corrosive water. For all systems, if the monitoring shows that more than 10% of samples taken from high-risk residences exceeds a Lead Action Level of 15 parts per billion (ppb) the water systems must undertake a series of actions. These include system-wide corrosion control treatment, source water monitoring, and ultimately lead service line replacement. There is no safe level of lead in humans, so the 15 ppb action level for public water systems is not a health-based standard. Rather it is "a trigger for treatment rather than an exposure level."¹⁵ The Lead and Copper Rule does not directly apply to schools or childcare facilities, unless they are labeled a public water system.



¹⁴ Hitlin, P., Olmstead, K., and Toor, S. 2017. Public Comments to the Federal Communications Commission About Net Neutrality Contain Many Inaccuracies and Duplicates. Pew Research Center. Accessed at: <https://www.pewresearch.org/internet/2017/11/29/public-comments-to-the-federal-communications-commission-about-net-neutrality-contain-many-inaccuracies-and-duplicates/>

¹⁵ USEPA. 2006. 3Ts for Reducing Lead in Drinking Water in School: Revised Technical Guidance. Washington, D.C: Author.

PROPOSED LEAD AND COPPER RULE REVISIONS

The [proposed revisions to the Lead and Copper Rule](#) maintains the action level of 15 ppb for lead in 10% of the tested samples. However, it adds a new “trigger level” of 10 ppb, when water utilities are required to take action. Those required actions include consulting with their state agencies on planning and monitoring, and implementation of corrosion control treatment. Any systems with lead service lines must develop a lead service line removal plan. Additionally, if a private property owner chooses to replace their portion of a lead service line, utilities are required to replace the utility-owned section of the same lead service line within 45 days. In a major change, the proposed rule requires utilities to replace 3% of the lead service lines annually for two years if the action level of 15 ppb is attained during regular testing. On paper, this is a significant rollback of the current regulations which require 7% annual lead service line replacement. However, the current rule’s requirement is rarely implemented and testing is less extensive, so some experts believe this change may prove a positive one once the rule is finalized.¹⁶

Major positive changes in the revision include a requirement for all utilities to develop and make publicly available and annually update their lead service line inventory, notify customers within 24 hours if the lead action level is reached, and test for lead in 20% of K-12 schools and daycare centers annually. The draft requires utilities to notify customers if the presence of lead pipes in their water supply is unknown, and the revisions provide further flexibility to water systems serving under 10,000 people to decide their lead mitigation approach.



**THE EPA RECEIVED NEARLY 80,000 COMMENTS
AT THE END OF COMMENT PERIOD.**

PUBLIC COMMENT PROCESS

After its publication in the Federal Register on November 13, 2019, EPA accepted public comments on the rule over a 90-day period, having provided a 30-day extension to the usual 60-day comment period after numerous requests from utilities, trade groups, and community advocates. The EPA received nearly 80,000 comments at the end of comment period. Of these, slightly fewer than 700 are unique and substantial, and are made publicly available on regulations.gov website. The EPA is expected to release the final rule to the public in October 2020.

¹⁶ Neltner, T. 2020. Personal communication. January 2020.

APPLICATION OF NATURAL LANGUAGE PROCESSING TO PUBLIC COMMENT ANALYSIS

Our aim was to use natural language processing techniques to develop insights into the qualitative characteristics of submissions to the LCR revision call for comments. Specifically, we sought to (1) acquire all available submission data from the open comment period, (2) perform topic modeling analysis on textual data in order to discover key repeated thematic patterns in submissions, and (3) perform sentiment analysis on textual data in order to characterize emotive language in submissions. Furthermore, we sought to explore relationships between the textual data and the type of agency that submitted the document. We focused our analysis on the following categorization of submitting agencies: elected official, equipment manufacturer/private consultant, member of the public, public health/advocacy group, schools, state agency, trade association, utility, or other. This classification was arrived at by the authors and differs from the categorization provided by regulations.gov.

EXTRACTING INFORMATION FROM SUBJECTIVE PUBLIC COMMENTS

1

Acquire all available submission data from the open comment period

2

Perform topic modeling analysis on textual data in order to discover key repeated thematic patterns in submissions.

3

Perform sentiment analysis on textual data in order to characterize emotive language in submissions.

We analyzed 1,546 public submission records, and delineated between two sources of language data: 505 unique comments (free text) and 382 unique attached documents (supporting document). We obtained key observations using two natural language processing techniques: (1) topic analysis (using the Latent Dirichlet Allocation model) and (2) sentiment analysis. Using these techniques, we identified several major findings.

COMMENTS VERSUS ATTACHMENTS

Content source (“comment” vs. “attachment”) is a greater differentiator than initial submitter categorizations developed by regulations.gov (labeled “document sub-type”). The language used in comments and attachments has different lengths, word frequencies, latent thematic patterns, and emotive content. “Comment” language has more emotional content while language in “attachments” makes appeals to reason via presentation of evidence and use of formal language. We believe that this reflects major differences in authorship and therefore stakeholder perspective. This is supported by the observation that the language used in comments and attachments can be used to classify the type of organization that made the submission, which suggests differences in language usage by the general public, companies/organizations, and government. In addition, we find that different proportions of “author” categories among comments and attached documents (see **Figure 3**).

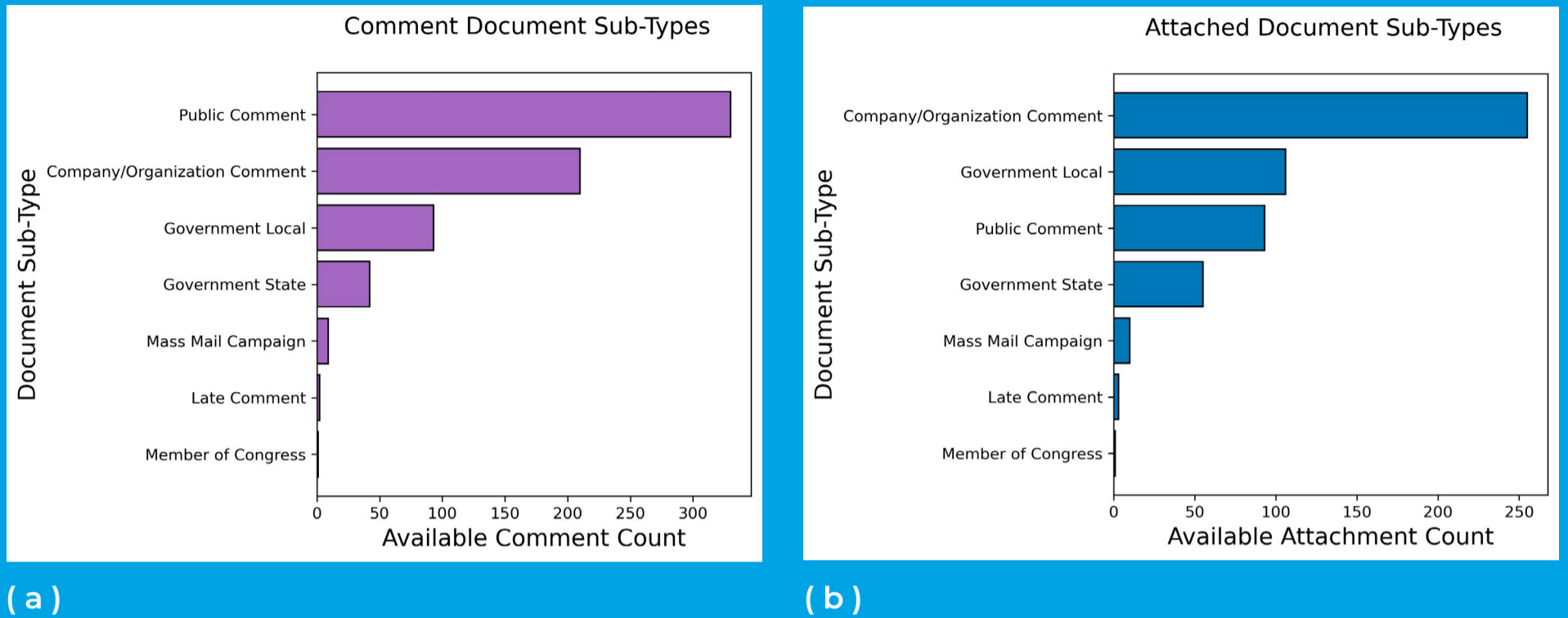


FIGURE 3

The counts of unique available (a) comments and (b) attachments organized by their submitting “documents sub-type”, as provided by the regulations.gov API.

One hundred and forty-three attachments were excluded due to duplication or inability to machine-read. A comment submitted by BlueConduit was excluded to avoid any conflict of interest. The remaining 239 attachments were further analyzed. Nearly a third of the attachments were submitted by water utilities (**Figure 4**).

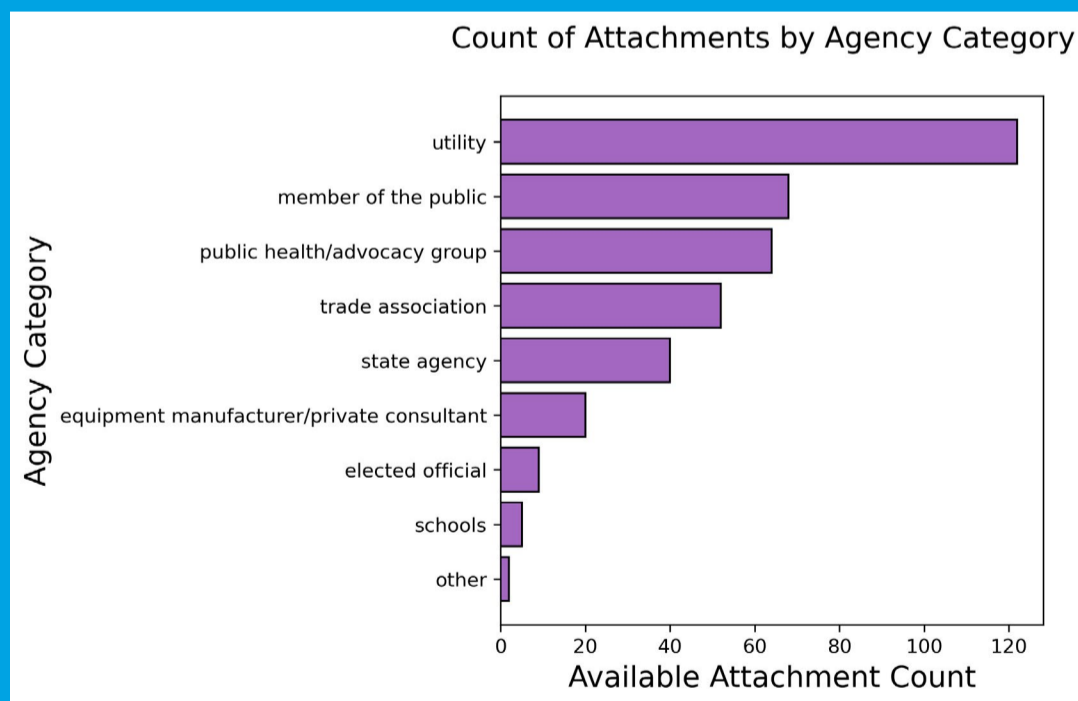


FIGURE 4

Count of attachments submitted by various commenter. Utilities were the largest group, followed by the public and public health/advocacy groups.

TOPICS COVERED

Thematic patterns differ by content source. Topic analysis found evidence that the topics differ subtly depending on their content source (“comment” vs “attachment”). We present the results of two topic models; one for comments (**Table 1a**) and one for attachments (**Table 1b**). Although there are notable differences between the two content sources, there is no evidence that these topic mixtures can be used to classify or predict the type of organization that made the submission. This may suggest that thematic coverage is not dependent on authorship (general public, company/organizations, and government). In other words, there is clear evidence that different types of commenters discuss the same thematic material as one another.

(a)

Topic	Comment Keywords
1	lines, revisions, problem, also, public, attached comments, levels, less, pipes, please see
2	please see, attached, proposed, schools, also, revisions, problem, children, would, lines
3	action, lines, please see, contaminated, children, safe, utilities, comments, proposed, new
4	lines, proposed, replace, elevated, time, action, public, epa s, would, contaminated
5	see attached, schools, epa, children, safe, also, exposure, proposed, comments, san francisco
6	proposed, quality, regulations, epa, s, also, people, public health, communities, attached comments
7	lines, epa, public, well, safe, also, works association, health, sources, like
8	epa, lines, children, levels, testing, exposure, public, action, sampling, proposed
9	schools, san francisco, exposure, safe, epa, lines, levels, children, levels, proposal
10	lines, epa, testing, proposal, children, public, new, would, issue, poisoning

(b)

Topic	Comment Keywords
1	lines, revisions, problem, also, public, attached comments, levels, less, pipes, please see
2	please see, attached, proposed, schools, also, revisions, problem, children, would, lines
3	action, lines, please see, contaminated, children, safe, utilities, comments, proposed, new
4	lines, proposed, replace, elevated, time, action, public, epa s, would, contaminated
5	see attached, schools, epa, children, safe, also, exposure, proposed, comments, san francisco
6	proposed, quality, regulations, epa, s, also, people, public health, communities, attached comments
7	lines, epa, public, well, safe, also, works association, health, sources, like
8	epa, lines, children, levels, testing, exposure, public, action, sampling, proposed
9	schools, san francisco, exposure, safe, epa, lines, levels, children, levels, proposal
10	lines, epa, testing, proposal, children, public, new, would, issue, poisoning

TABLE 1

The counts of unique available (a) comments and (b) attachments organized by their submitting “documents sub-type”, as provided by the regulations.gov API.

AGENCY PRIORITIES

Topic analysis found detectable differences between different commenter types, but that these differences do not always correspond to differences of opinion (**Figure 5**). In general, language use is highly conserved since the focus of the documents is very specific. Some observed differences are likely indicative of authorship styles, rather than differing perspectives. For example, members of the public and advocacy organizations tend to discuss the same topics, but the language used by members of the public is less formal and can be more clearly identified using automated approaches. Although we did not test this in a statistical way, in general it appears possible to identify the letters submitted by members of the public but that were written by advocacy organizations. However, we found some evidence that automated systems could confirm what is obvious to a human reader: that different commenters have differing priorities. For instance, after removing the 30 most common phrases, advocacy groups and members of the public were found to more frequently use keywords like “exposure”, “children”, “levels”, whereas utilities and trade association used keywords such as “customer”, “corrosion control”, and “treatment.”

TOPIC ANALYSIS FOUND DETECTABLE DIFFERENCES BETWEEN DIFFERENT COMMENTER TYPES, BUT THAT THESE DIFFERENCES DO NOT ALWAYS CORRESPOND TO DIFFERENCES OF OPINION.

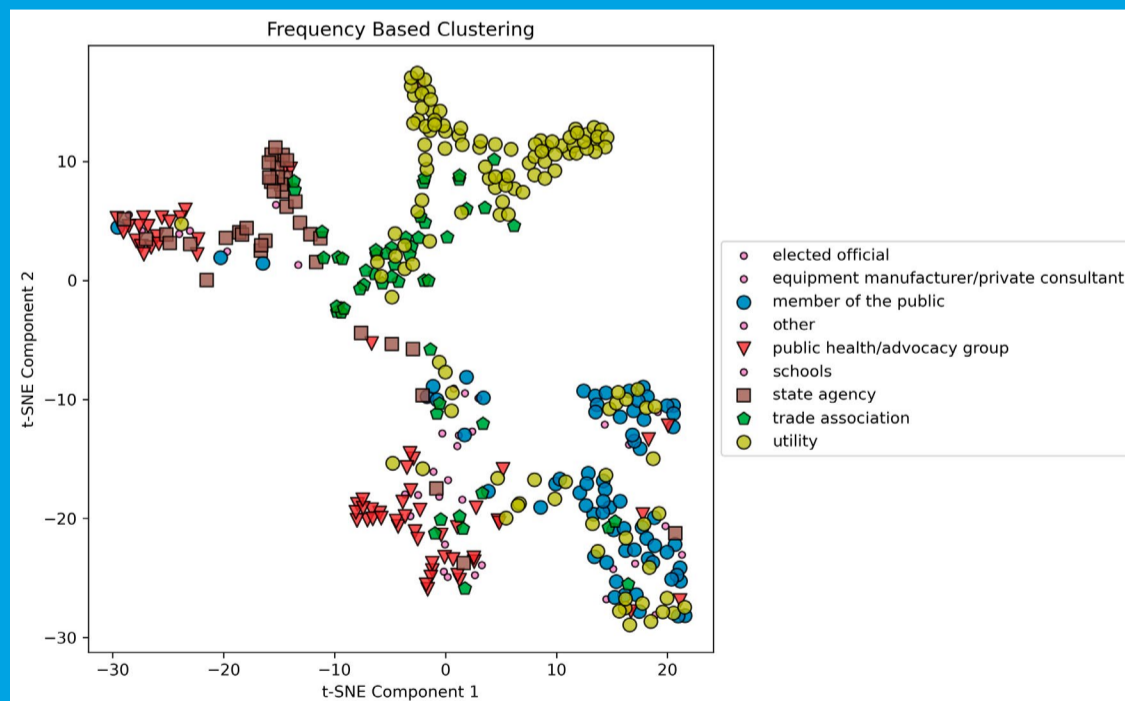


FIGURE 5

Attached documents clustered by their top-10 most frequent phrases up to 3 words long. Strong separation is evident between agency types and groups are largely dominated by a single agency type, but there are some documents distributed throughout the plane.

CHILDREN'S SAFETY

Concern for children's safety is a key theme. Machine learning tools found strong evidence that "children's safety" is a central, repeated thematic pattern in both comments and attachments. See examples of particularly strongly positive and negative sentences extracted from the comments that related to child safety:

Positive statement: *I trust that my remarks can be taken constructively and may help lead to a safer and healthier environment for children, our most precious "possessions," and only hope for the future!*

Negative statement: *If a child consumed water from those fountains, there is a higher percentage that the child will suffer some consequences such as criminal behavior, reading problems, attention deficit hyperactivity disorder, school failure, delinquency, and antisocial behavior.*



MACHINE LEARNING TOOLS FOUND STRONG EVIDENCE THAT "CHILDREN'S SAFETY" IS A CENTRAL, REPEATED THEMATIC PATTERN IN BOTH COMMENTS AND ATTACHMENTS.

REPLACEMENT OF LEAD LINES

While calls to enhanced testing and sampling of drinking water are frequent, machine learning found that service line replacement was a more frequent and detectable theme of public input. This theme is central to “attachments”, and present, though to a lesser extent in comments (see **Table 1**). Sentiment analysis suggests that different groups of commenters - especially utilities and members of the public - express similar sentiments on lead service line replacement (**Figure 6**). The analysis detected more highly positive documents than highly negative documents for most groups of commenters. The median is closer to the zero-point for elected officials, indicating approximate symmetrical sentiment with respect to service line replacement. Similarly, there were central, repeated mentions of “corrosion control” and “action” in “attachments.”

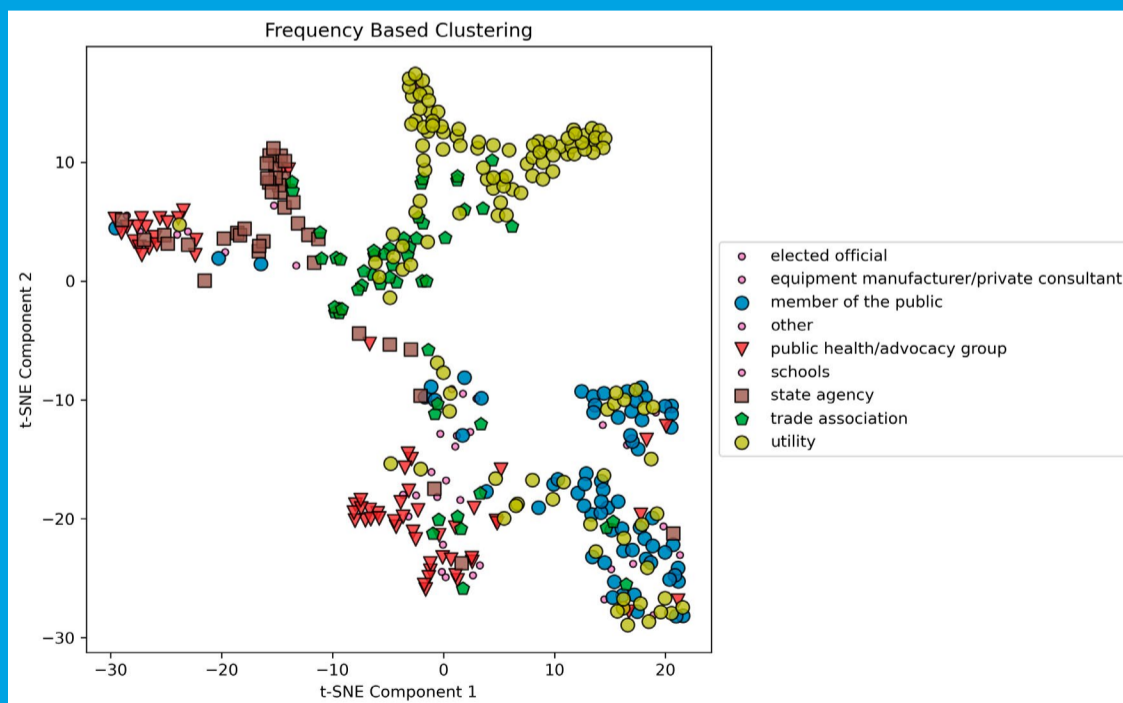
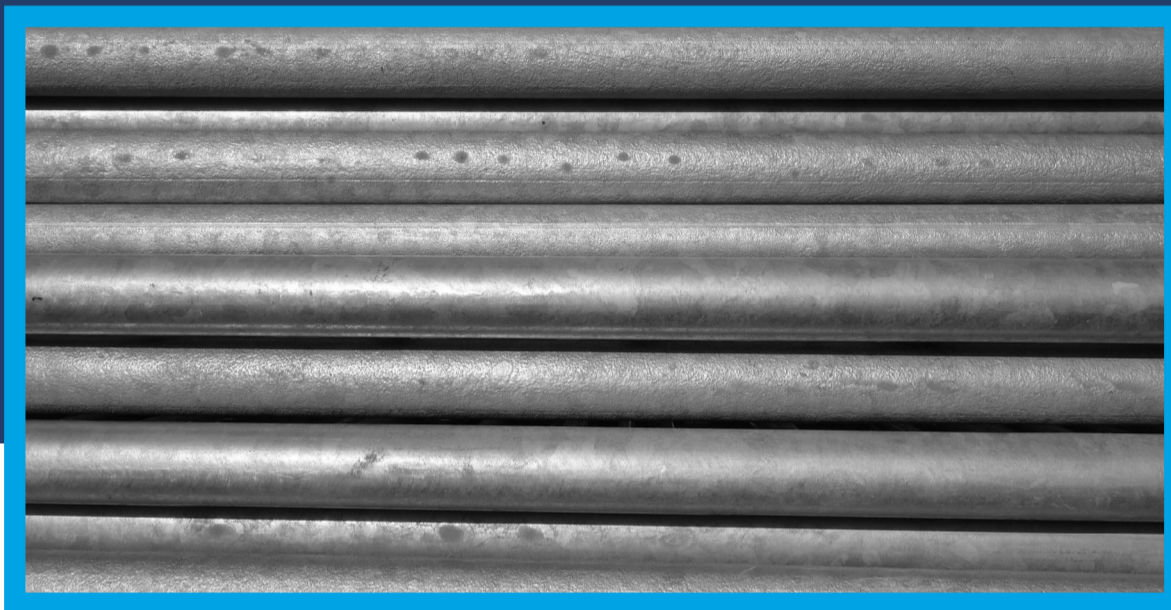


FIGURE 6

Distribution of average sentiment scores for documents containing the word 'replace', 'replacement', or 'LSLR'.

TRIGGER LEVEL

The EPA's introduction of a trigger level (10 ppb) in addition to the action level (15 ppb) resulted in several statements referring to this term. A boxplot of the sentiment score across all documents within a group suggests the ranges (box size and whisker length) vary across different groups (**Figure 7**). The outlier ranges are compact, indicating a relatively consistent sentiment with respect to trigger levels. Interestingly the analysis found more sentiment variability from groups such as utilities than from the general public, which makes sense given the potential impact of the proposed changes. The median sentence sentiment is positive for all agency types, but that 'elected officials' IQR straddles the zero-point, indicating a large proportion of these documents expressed mildly negative sentiment for sentences containing a word related to 'trigger level.'

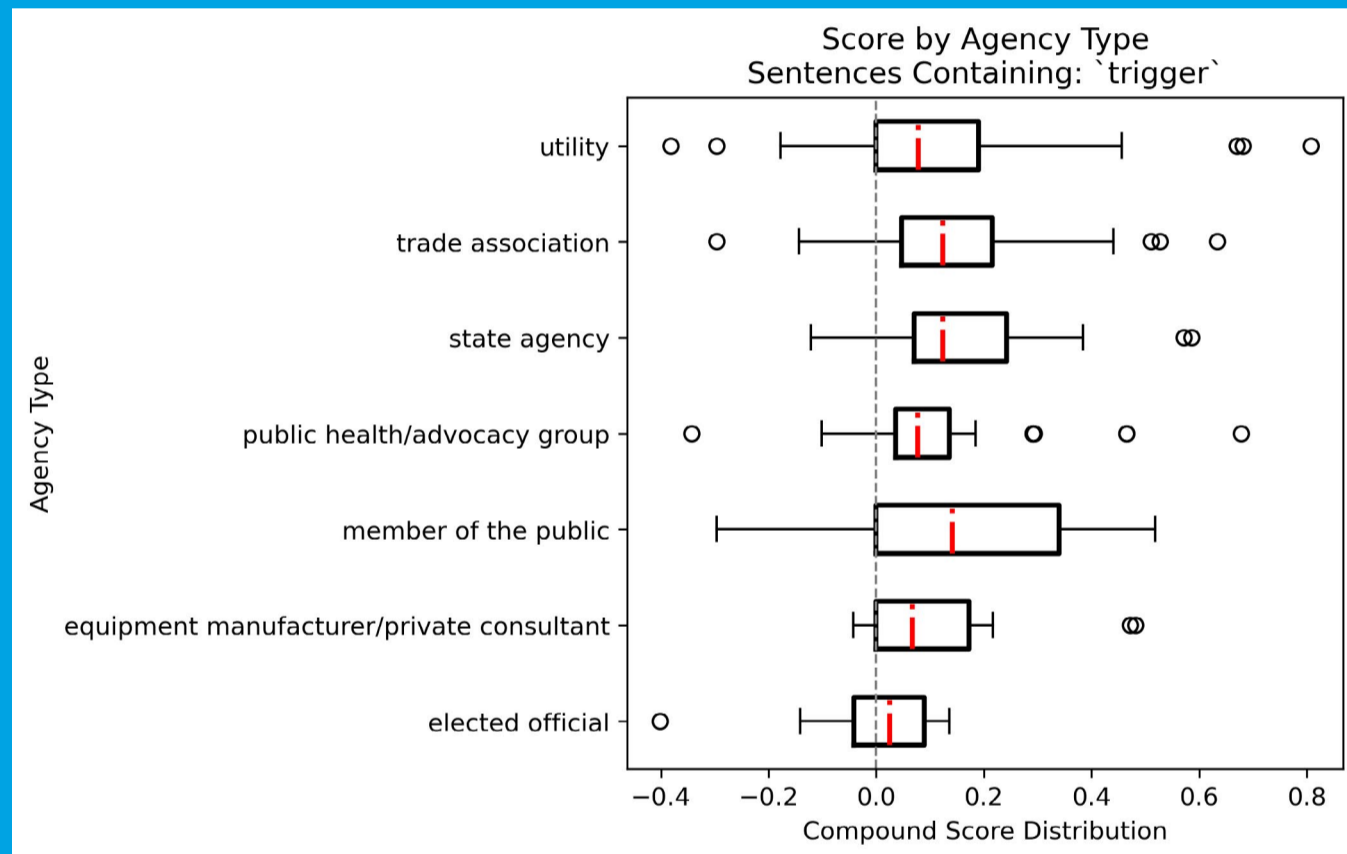


FIGURE 7

Distribution of average sentiment scores for documents containing the word 'trigger', 'trigger level', 'action level', or 'threshold.'

COST CONSIDERATIONS

The very little mention of costs and funding found in the documents conformed to expected hypotheses: utilities described the proposed rule (or parts thereof) as “cost prohibitive”, while public health groups argued that the “public health benefit of lead reduction in water exceeds the cost of doing so.” The analysis found little to no evidence that financial themes were central in both comments and attachments. Additionally, there was no evidence of negative sentiments related to “unrealistic” financial demands, suggesting that the federal government should not exaggerate cost considerations in its final rule, compared to the draft. The general absence of comments on costs or funding went against our expectations.

THERE WAS NO EVIDENCE OF NEGATIVE SENTIMENTS RELATED TO “UNREALISTIC” FINANCIAL DEMANDS, SUGGESTING THAT THE FEDERAL GOVERNMENT SHOULD NOT EXAGGERATE COST CONSIDERATIONS IN ITS FINAL RULE, COMPARED TO THE DRAFT.

OVERALL CONTENT ANALYSIS

Sentiment analysis on text included in comments and attachments presents compelling evidence that the linguistic content of comments and attachments are qualitatively different from one another. Specifically, we find sentimental content (positive/negative phrases) to be twice as much in comments compared to attachments. This reinforces our observation based on reading comments that comment text is more likely to use emotion, whereas attachments appeal to logic/reason. We find that the average compound polarity scores for attachments are slightly positive (mean: 0.028, 0.95CI 0.028 - 0.029) (**Figure 8**). However, on average, attachments were dominated by ‘neutral’ phrases (mean neutral proportion: 0.936, 0.95CI 0.935 - 0.936).

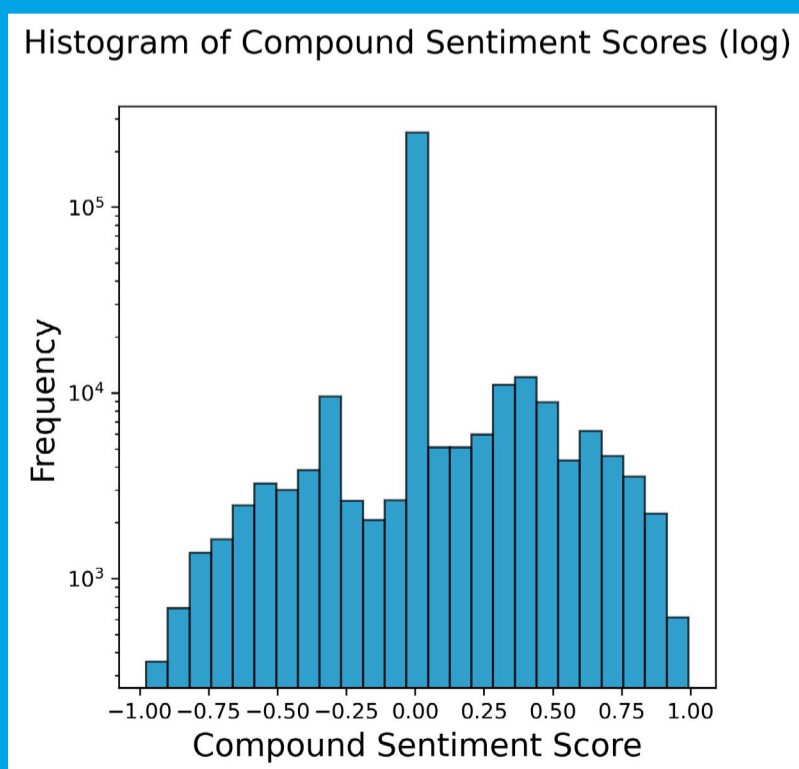


FIGURE 8

Distribution of compound sentiment scores on text within attachments.

SUMMARY

Our analysis found detectable differences between different agency types, but that these differences do not always correspond to differences of opinion. Some observed differences are likely indicative of authorship styles, rather than differing perspectives. Sentiment analysis found significant differences in the linguistic content of comments and attachments; the comment text appealed to emotion, whereas attachments appealed to logic/reason. However, we found the documents to be largely neutral in tone, with only a slightly positive sentiment. Further analysis of documents containing specific keywords related to “replacement”, “trigger level” and “children” confirmed the same. In summary, we found evidence that submissions differ in how they present their agenda but are reasonably consistent about what that agenda is. Natural Language Processing is an exciting tool to assess differences in sentiment and content across different authors and commenters in a large set of text, but currently available techniques are not sufficient to parse language and tone to provide decisive conclusions about the authors’ intent and judgment. Lack of rigorous categorization of ‘submitter’ in the regulation.gov API metadata hindered faster and accurate processing of comments. We nevertheless find it a promising area of study and constant development.

APPLICATIONS FOR FURTHER RESEARCH

Using more optimized tools, we believe automated sorting of comments for high-volume regulations will soon become necessary to ensure agencies’ response to comments are more accurate and less subject to bias. Although challenging to apply, analyses like ours and that of FiscalNote show that agencies like the GAO, Congressional Research Service, and White House Office of Management and Budget will gain tremendously by automating the detection of product and process patterns across the federal government and over time. We expect that rapid development of machine learning technologies will soon make them invaluable for individual rule review and as a tool for courts to determine whether agencies have adequately addressed major elements of public comment and interpreted those comments accurately.

