Approved for Public Release; Distribution Unlimited. Public Release Case Number 22-2533

AI-Based Text Analysis for Evaluating Food Waste Policies

John A. Aitken,[1] Denali W. Rao, Balca Alaybek, Amber Sprenger, Grace Mika, Rob Hartman, Laura Leets

The MITRE Corporation

[1]jaitken@mitre.org

Abstract

Food waste is a major contributor to climate change, making the reduction of food waste one of the most important strategies to preserve threatened ecosystems and increase economic benefits. To evaluate the impact of food waste policies in this arena and provide actionable guidance to policymakers, we conducted an AI-based text analysis of food waste policy provisions. Specifically, we used unsupervised machine learning to a) identify commonalities across state policy texts, b) cluster states by shared policy text, and c) examine relationships between state cluster memberships and food waste. This approach generated state clusters but demonstrated very limited convergent validity with policy ratings provided by subject matter experts and no predictive validity with food waste. We discuss the potential of using supervised machine learning to analyze food waste policy text as a next step.

Keywords: food waste; date label policy; text analysis

Introduction

Food waste is one of the most significant drivers of climate change, constituting up to 10% of all greenhouse gas emissions, 14% of all water use, 18% of all cropland use, and 24% of all landfill content (Hall et al., 2009; Quested, Ingle, and Parry, 2013). The reduction of food waste is a key climate change strategy (Hawken, 2017), and is a challenge that depends on a host of actors across all steps of the supply chain. Strong federal and state policy is one of the most promising avenues for mitigating food waste and stimulating food recovery (Evans and Nagele, 2018). One key food waste policy in this area and which is well-represented at the state level is date label policy.

Date label policy applies to whether manufacturers must include labels on certain food (e.g., milk, meat) and whether the product may be sold past the date, and other requirements such as the use of specific terminology (e.g., "Best by," "Sell by," "Use by"). Currently, instead of a federally-enforced standard policy, there is a patchwork of date label policies across states that grants free reign to manufacturers and, in turn, creates confusion for consumers (Broad Leib and Pollans, 2019; Broad Leib et al., 2016). For example, consumers may incorrectly believe that food should be discarded once past its date, whereas some dates may only be indicators of quality rather than wholesomeness (Busetti, 2019). It has been suggested that states with more extensive date label policy (i.e., more requirements and/or restrictions) contribute to rather than reduce food waste (Lipinski et al., 2013; Povich, 2020). However, there is a paucity of empirical evidence regarding the impact of these policies on relevant outcomes.

Therefore, it is essential to empirically evaluate the extent to which state date label policies contribute to waste and impacts the environment. One challenge in accomplishing this goal is that these policies can be generally opaque due to legal jargon and require expert analysis to distill. For example, beyond reading and comprehending a state's date label policy, an expert must also be able to evaluate the strengths and weakness of the policy with respect to relevant policy at the federal level and in other states as well as with respect to the nuances of the issue at hand (i.e., knowing which policy features are more or less beneficial in combating climate change). In such an evaluation, subject matter experts may differ in the extent to which they agree about which policy features should be considered (i.e., which are most relevant to the efficacy of the policy) as well as how to judge qualitative aspects of the policy (e.g., strength, extensiveness), requiring prolonged discussions and recalibration. Such an effort is extensive and time-consuming, which is far from ideal given the significance of food waste in driving climate change and the urgency for policymakers to craft effective and relevant policies.

In this respect, text analysis with the aid of artificial intelligence (AI) represents a promising avenue of policy evaluation in the food waste and climate change domains. First, AI-based text analysis may be an extremely efficient tool in analyzing large quantities of policy text and rendering a list of key characteristics that differentiates one state's policy from another. This would greatly benefit domain experts and researchers in general by functioning as a powerful and flexible tool in many climate change policy areas (Short, McKenny, and Reid, 2018). For example, there may be a high degree of shared text between policy texts that may indicate similarities between state policies. Also, there may be unique features of certain policies differentiating them from others. Overall, then, legislative text reuse and analysis could serve as a window into the spread of political influence (Wilkerson, Smith, and Stramp, 2015).

Moreover, developing an analytical method that can summarize and evaluate climate change policy text may allow for non-experts to investigate and interpret this policy area. This enables a multidisciplinary approach to a typically complex legislative area, and such an approach is crucial given the scope of climate change and its causes (of interest to this paper, food waste) as well as the variety of domains (e.g., environmental science, political science, social science) and stakeholders (e.g., federal agencies, nonprofit organizations) involved in addressing these problems.

To that end, the natural language processing field combines AI and computational linguistic techniques and provides a variety of machine learning approaches (e.g., supervised, unsupervised) for text analysis. The rest of this paper summarizes our application of unsupervised machine learning to food waste policy text analysis. Specifically, we performed a text analysis of U.S. state food date label policies to derive state clusters that (1) meaningfully represented the content of shared policy text, (2) converged with human subject-matter expert ratings of policies, and (3) predicted food waste. The content, convergent, and predictive validation of such a method would contribute to impact analyses in not only date label policies but also other climate change policy areas.

Method

The data for this study included (1) date label policy texts from 50 U.S. states enacted prior to 2012 and (2) municipal solid waste (MSW; 22–24% of which is estimated to account for food waste across states, U.S. Environmental Protection Agency, 2022).

Policy Text Preparation for Text Analysis

We processed the policy text at two levels of analysis: entire provisions (i.e., sections with explicit citation labels) and individual clauses within legislative provisions (i.e., distinguished by line breaks and enumeration marks). We organized the data accordingly and removed duplicate provisions, which resulted in 113 distinct provisions and 1846 distinct clauses in the date label policy dataset. We then tokenized the text, using term frequency-inverse document frequency (TF-IDF) weighting to create token-frequency vectors. Finally, we discarded policy fragments with fewer than seven tokens as they were too short to be meaningful.

Policy Text Coding for Validation

Food waste policy subject matter experts (SMEs) from Harvard Law School, Food Law and Policy Clinic (HFLPC) manually coded policy texts to generate a ground-truth characterization of the state date label policies against which the text analysis-derived clusters could be validated. We used three manually coded variables for validation purposes: for a given food type, (1) whether a date label is required, (2) whether sale after label date is restricted, and (3) whether the policy required the use of specified date label terminology (e.g., "use by," "sell by," "best by"). We created three respective continuous variables (i.e., DateTotal, SaleTotal, and TermTotal) that indicated the number of food types for which a given date label policy was enacted in a given state.

AI-Based Text Analysis Plan

We took three steps in each policy text analysis. First, we applied a standard topic modeling algorithm, Gensim, to the provision token frequency vectors. Given our set of non-uniformly structured text data, topic modeling was the natural choice to begin processing and understanding the data. Although there were 8 distinct food types identified in our dataset as policy foci (Breads & Bakery, Dairy & Eggs, Dry Goods, Fresh Meat & Seafood, Frozen, Prepared Foods, Produce, and Ready-to-drink Beverages), it was important to represent the data with more topics than just these 8 to capture all possible fragments and more specific food types (e.g., Shellfish within the broader Fresh Meat & Seafood category). Additionally, while it is generally recommended to run the Gensim topic modeling algorithm with 300-500 topics, we determined that our dataset was unlikely to include as many distinct topics (Bradford, 2009). Preliminary experimentation revealed that extracting more than 150 topics yielded many overlapping topics, while extracting fewer than 80 topics from the dataset yielded topics that incorporated unrelated concepts into one. Accordingly, we specified the model to generate 100 topics. We represented the policy text fragments as proportions of the 100 topics and computed the cosine similarity between each pair of policy fragment topic vectors. We then applied a similarity threshold to select only the stronger relationships between policy fragments and generated a network graph to visualize the results, plotting fragments as nodes and the relationships between them as edges.

Second, we attempted to group the policy fragments by their semantic features. We used agglomerative clustering on the previously generated network graph to detect groups of similar policy fragments and color-coded the graph to reflect these clusters. Using a hierarchical clustering method like agglomerative clustering allowed us to leave the number of clusters to create unspecified and explore how many clusters "naturally" emerged from the data. The resulting policy fragment clusters from this step, since they were computed via the application of similarity metrics to topic modeling outputs, represented equivalence classes under topic similarity.

Third, we used the policy fragment clusters as features and described each state as a combination of the features it had. These descriptions took the form of feature vectors (similar to the topic feature vectors we saw earlier, but with one per state instead of per policy fragment). For example, when a state had one policy fragment that fell under a dairy labeling cluster, had two that fell under the shellfish requirements cluster, and had no policy about pork (meaning no membership in a pork requirements cluster), each of these cluster memberships as well as non-memberships were incorporated in the state's feature vector. We then took the cosine similarity of these state feature vectors and applied another similarity threshold to select only strong relationships between states. From the resulting filtered state similarity matrix, we generated a network graph using states as the nodes in the graph and reflecting the strength of the similarity between them in the lengths of the edges. Finally, we detected clusters of states in the graph and color-coded those clusters. We expected these state clusters to consist of states that were similar to each other with respect to their food date label policy content.

Results

Text Reuse Analysis

We began by examining verbatim text reuse by employing common subsequence analysis to compute all common subsequences of at least 6 words between each pair of states. We then used three analytical techniques: (1) extracting the length of the longest common subsequence between every pair of states, (2) computing the number of common subsequences shared by each pair of states, and (3) extracting any subsequences greater than 6 words long that were common to more than two states. However, this approach failed to identify identical provisions at the section level nor verbatim duplication of meaningful policy expressions within provisions. Therefore, we instead shifted our approach to searching for *similar* text between policies at the level of the provision and of individual clauses, hypothesizing that states with similar policy text share common policy objectives and may be clustered as such.

Date Label Policy Text Analysis

Beginning with date label policy text at the provision-level, we used the topic modeling algorithm to generate 100 topics. Figure 1 shows six of the twelve most significant topics' ten most strongly weighted tokens compared to the frequency of those tokens in the entirety of the text. Some topics were more clearly interpreted than others: Topic 0 (shellfish, tag, dealer, molluscan, shucked, shellstock, etc.) clearly revolved around shellfish and how they should be caught and processed; Topic 2 (egg, milk, carton, pack, size, inch, etc.) seemed to be about specifically egg cartons in contrast to milk cartons; and Topic 10 (mean, sandwich, expiration, prewrapped, open, vendor, etc.) seemed to be about prewrapped sandwiches and their expiration dates. Each provision was represented as a vector combination of the 100 topics generated by our topic modeling algorithm.

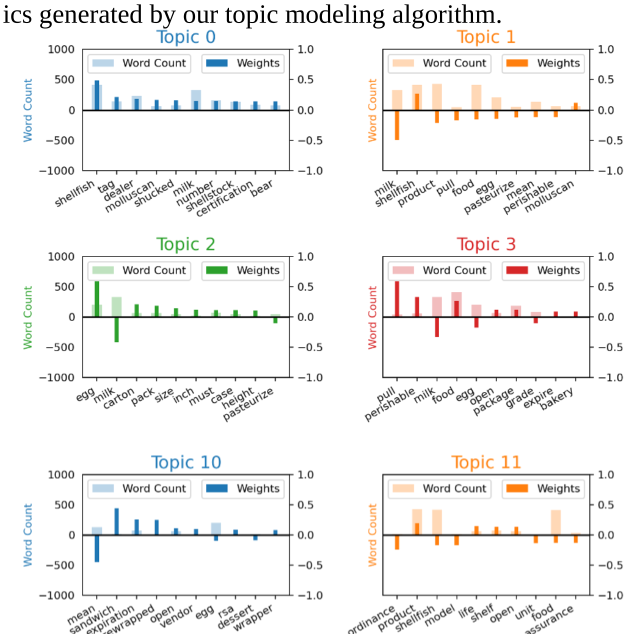

Figure 1. The ten most strongly weighted tokens from some of the top twelve most significant topics (Provision-Level Date Labeling Analysis)

We computed the pairwise cosine similarity between provisions based on this representation and applied a similarity threshold of 0.6 (discarding any values below the similarity threshold). Agglomerative clustering (with distance threshold setting of 1.5) identified 20 clusters of provisions. We used these clusters as features and represented the states as vectors of length 20 denoting which clusters their provisions fell into. Most of the clusters revolved around a certain food type (milk, shellfish, prewrapped sandwiches) while a few clusters were more general (pull dates, misbranding), so the number of features a state had was often a reflection of how many different food types that state's date label policies addressed. The number of features a state had was also partly a reflection of how many date-label-related provisions a state had in total.

Finally, we clustered the states themselves based on the feature clusters (see Figure 2). We created a binary state feature matrix, took the cosine similarity of the matrix, and discarded all values less than 0.5. Greedy modularity maximization yielded 8 clusters of states, with a modularity score of 0.66.

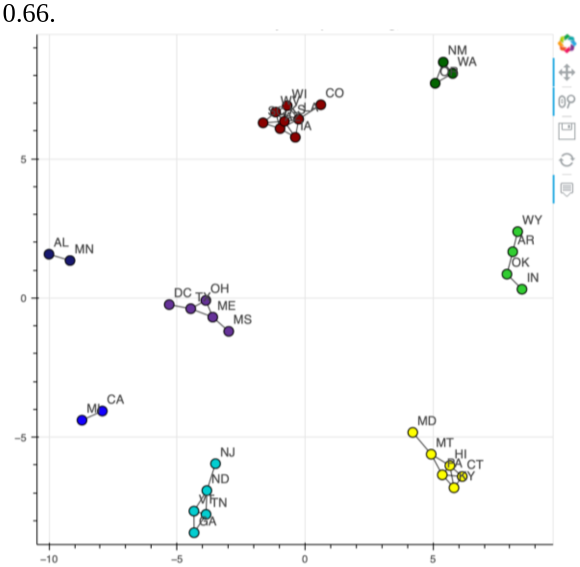

Figure 2. Clause-Level Date Labeling Analysis: State Clusters

Turning now to the clause-level analyses, many of the same significant tokens emerged as in the provision-level analysis, implying that the same tokens that were significant within a whole provision are still the most significant when the text is broken into smaller segments. On average, however, it was harder to ascertain what these topics were about. Also notable was that the most significant tokens in these topics were less strongly weighted than the most significant tokens in the provision-level topics. It is possible that the fragmenting of the text to the clause level also split up important or key phrases, so that individually each clause had fewer key phrases signaling its meaning.

Agglomerative clustering (with a distance threshold setting of 4) detected 72 clusters of clauses. These clusters, as may be expected, were more homogenous than the clusters found at the provision-level analysis, both because the unit of text was smaller and because there were more clusters for them to separate into. Greedy modularity maximization detected 6 clusters of states, with a modularity score of 0.69.

Validation Analyses

For convergent validation (i.e., examining the relationships between states' cluster memberships and SME-coded policy variables) and predictive validation (i.e., examining the relationships between states' cluster memberships and food waste), we first computed two continuous state topic count variables (i.e., number of topics within which a given state fell under). One of the continuous variables was based on provisions and the other was based on clauses. Also, in these variables, we included states that did not have any date label policies, which received a value of zero. We computed the Kendall's Tau correlations of the state topic count variables with the SME coded policy variables and the outcome variable (MSW), and found weak relationships ($\tau_b = 0.19$, $p = .079$ for provisions; $\tau_b = 0.15$, $p = .133$ for clauses). Moreover, after excluding the states with no date label policies from the topic count variable, topic count and DateTotal variables, this correlation remained nonsignificant.

In addition, we conducted a series of chi-squared difference tests to examine whether state clusters were related to SME-coded policy variables and MSW. These analyses returned nonsignificant results no matter states with no date label policies included.

Discussion

Our findings generally suggested that the unsupervised machine learning approach for text analysis was able to cluster food waste policy fragments and states based on similar features that emerged through the text, but the results demonstrated very limited convergent validity with those generated by SME coding and no predictive validity with the food waste outcome. Our work in progress involves validating a supervised machine learning approach to analyze policies relevant to food waste and climate change.

Additionally, future work could perform more pre-processing of the policy text and employ more sophisticated natural language processing (NLP) models. Although we started our analyses by preparing our text with standard and widely used text cleaning methods, legal text often contains additional levels of complexity (e.g., enumerations, hyper-specific abbreviations, particularly formal phrasings) compared to the type of text that our methods are commonly designed for and used on (e.g., social media posts, Wikipedia articles). Therefore, our dataset would likely benefit from additional processing that is more appropriate for policy text. One potential direction is to use a tool with pre-trained word vectors such as GloVe (Pennington, Socher, and Manning, 2014). Although it would be more computationally expensive, GloVe's incorporation of linguistic and semantic similarity between words might be useful. For example, equating the words "shellfish" and "mollusks" might illuminate some previously hidden policy similarities in our dataset. Another potential tool is LEGAL-BERT (Chalkidis et.al., 2020). The authors of LEGAL-BERT faced the same issue we note above—that the usefulness of standard preprocessing tools may not generalize to legal text. We could leverage their conclusion (i.e., pre-training BERT models on legal text improves performance) as well as their publicly released pre-trained models to improve our analyses.

Moreover, the present work considered only policy text related to food waste, which is an important arena in the climate change discussion (Hall et al., 2009; Quested, Ingle, and Parry, 2013), but future research may also consider legislative policy that is tied to other areas with a negative environmental impact. We examined food waste as an initial investigation and test of this methodology, and there is clear potential for the examination of not only other policy texts related to food waste (e.g., liability protection, tax incentives, etc.; Broad Leib et al., 2020) but related to sustainable fishing (Worm et al., 2006) and energy use (Hawken, 2017). These other areas are of obvious relevance to climate change and may include legislative policy that is amenable to such analysis.

Finally, a limitation of the current work is the reliance on MSW as a proxy variable of food waste. Despite the likelihood that food waste exhibits significant convergence with MSW, the limited predictive validity that we found in the present work may be due to the MSW variable being a broader measure that includes non-food related waste. Thus, future work may evaluate food waste policies using a more proximal or narrowly defined outcome variable. In general, careful consideration of an appropriate outcome or indicator variable is especially important in empirical evaluations of legislative policy.

References

Bradford, R. 2008. An empirical study of required dimensionality for large-scale latent semantic indexing applications. CIKM '08: 153-162.

Broad Leib, B.; Rice, C.; Neff, R.; Spiker, M.; Schklair, A.; and Greenberg, S. 2016. Consumer Perceptions of Date Labels: National Survey. *Safety*, 23(54): 1-4.

Broad Leib, E., and Pollans, M. J. 2019. The New Food Safety. *California Law Review*, 107: 1173-1248.

Broad Leib, E.; Ardura, A.; Fink, B.; Hartman, M.; Giguere, M.; and Spiegler, R. 2020. United States Legal Guide: Food Donation Law and Policy. *The Harvard Law School Food Law and Policy Clinic: Cambridge, MA, USA*. Retrieved from: https://chlpi.org/wp-content/uploads/2013/12/USA-Legal-Guide-2020.pdf

Busetti, S. 2019. A Theory-Based Evaluation of Food Waste policy: Evidence from Italy. *Food Policy*, 88: 101749. doi.org/10.1016/j.foodpol.2019.101749.

Chalkidis, I.; Fergadiotis, M.; Malakasiotis, P.; Aletras, N.; and Androutsopoulos, I. 2020. LEGAL-BERT: The Muppets straight out of Law School. *In Findings of Empirical Methods in Natural Language Processing (EMNLP 2020)* https://aclanthology.org/2020.findings-emnlp.261

Evans, A. I.; and Nagele, R. M. 2018. A Lot to Digest: Advancing Food Waste Policy in the United States. *Natural Resources Journal*, 58(1): 177-214. jstor.org/stable/26394778.

Hall, K. D.; Guo, J.; Dore, M.; and Chow, C. C. 2009. The Progressive Increase of Food Waste in America and Its Environmental Impact. *PloS one*, 4(11): e7940. doi.org/10.1371/journal.pone.0007940.

Hawken, P. 2017. *Drawdown: The Most Comprehensive Plan Ever Proposed to Reverse Global Warming.* New York: Penguin Books.

Lipinski, B.; Hanson, C.; Waite, R.; Searchinger, T.; and Lomax, J. 2013. *Reducing Food Loss and Waste.* Washington: World Resources Institute.

Pennington, J.; Socher, R.; and Manning, C. D. 2014. GloVe: Global Vectors for Word Representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics.

Povich, E. S. 2019. Food Waste is a Major Problem. Confusing Date Labels are Making It Worse. *Stateline*.

Quested, T.; Ingle, R.; and Parry, A. 2013. *Household Food and Drink Waste in the United Kingdom 2012.* Banbury: WRAP.

Short, J. C.; McKenny, A. F.; and Reid, S. W. 2018. More than Words? Computer-Aided Text Analysis in Organizational Behavior and Psychology Research. *Annual Review of Organizational Psychology and Organizational Behavior*, 5: 415-435. doi.org/10.1146/annurev-orgpsych-032117-104622.

U.S. Environmental Protection Agency. 2022. *Greenhouse Gas Reporting Program*. Retrieved from https://www.epa.gov/ghgreporting.

Wilkerson, J.; Smith, D.; and Stramp, N. 2015. Tracing the Flow of Policy Ideas in Legislatures: A Text Reuse Approach. *American Journal of Political Science*, 59(4): 943-956. doi.org/10.1111/ajps.12175.

Worm, B.; Barbier, E. B.; Beaumont, N.; Duffy, J. E.; Folke, C.; Halpern, B. S.; ... and Watson, R. 2006. Impacts of Biodiversity Loss on Ocean Ecosystem Services. *Science*, *314*(5800): 787-790. doi.org/10.1126/science.1132294

Acknowledgments

This research was funded by the MITRE Independent Research and Development Program.

The authors thank Charles A. Worrell for his constructive feedback on earlier drafts of this paper.