Exploring the Role of Sponsoring Agencies in Shaping the MUTCD Using Supervised and Unsupervised Text Mining

Boniphace Kutela, Ph.D., P.E.[1]; Hellen Shita, S.M.ASCE[2]; Subasish Das, Ph.D.[3]; Lucy Kapaya[4]; and Elizabeth Tarimo[5]

Abstract: Preparing the *Manual on Uniform Traffic Control Devices* (MUTCD) involves gathering inputs from various public and the National Committee on Uniform Traffic Control Devices (NCUTCD) sponsoring agencies. The NCUTCD assists in developing standards, guidelines, and warrants for traffic control devices and practices employed to regulate, warn, and guide traffic flow on roadways. This organization advises the Federal Highway Administration (FHWA) and other relevant agencies on suggested updates and interpretations to the MUTCD and other nationally recognized standards. Examples of such sponsoring agencies include AASHTO, ASCE, Institute of Transportation Engineers (ITE), American Highway Users Alliance (AHUA), National Association of City Transportation Officials (NACTO), National Association of County Engineers (NACE), International Bridge, Tunnel and Turnpike Association (IBTTA), and American Traffic Safety Services Association (ATSSA). Other national and regional agencies have also been invited to provide feedback on the draft manual. Although all comments are considered valuable, those from federal agencies may hold more weight due to their expertise in specific subject matters. This study utilized text-mining techniques to analyze the comments from the sponsoring agencies, aiming to identify commonalities and distinct features in their feedback. The main question addressed was whether each agency effectively represented its interests and if there were shared interests among them. The study also explored the suggestions, questions, and recommendations raised by each agency. The findings revealed that some concerns were shared among agencies, although in most cases, they represented individual interests. Specifically, some agencies shared a common interest in speed limits, whereas others had similar topics concerning signs and markings. Other shared interests are bike lanes, 85th percentile speed, and pedestrian- and bus lane–related topics. This study's insights into the interests and concerns of various MUTCD sponsoring agencies can aid in creating more effective, engaging, collaborative, and well-rounded technical documents. As a limitation, this study only used the current MUTCD revision comments, and hence the results reflect only recent concerns from the agencies. **DOI: 10.1061/JTEPBS.TEENG-8395.** © *2024 American Society of Civil Engineers.*

Practical Applications: The Federal Highway Administration collects comments from both the general public and NCUTCD sponsoring agencies for modifying the MUTCD. This study used the comments from the NCUTCD sponsoring agencies submitted by May 2021 and conducted supervised and unsupervised text mining analysis to understand any available patterns, differences, or commonalities in the agencies' concerns in shaping the MUTCD. The findings indicate the presence of various shared interests among the agencies and some specific features that were specific to the agencies. In addition, it was observed that the majority of the agencies have a great concern for the safety and well-being of society by showing a common interest in safety influencing factors such as speed, road signs, and the definition and use of the 85th percentile speed. This understanding can foster greater collaboration among the agencies, leading to more well-rounded and comprehensive guidelines.

Author keywords: *Manual on Uniform Traffic Control Devices* (MUTCD); Sponsoring agencies; ASCE; AASHTO; American Highway Users Alliance (AHUA); National Association of City Transportation Officials (NACTO); National Committee on Uniform Traffic Control Devices (NCUTCD).

Introduction

The *Manual on Uniform Traffic Control Devices* (MUTCD) for streets and highways is a comprehensive set of standards used nationwide to install and maintain traffic control devices on public streets, highways, and bikeways (FHWA 2009). Maintained by the Federal Highway Administration (FHWA), it provides guidelines and specifications for the design, application, and placement of all traffic signs, signals, markings, and other devices used to regulate, warn, and guide road users (Jewell 2017). The MUTCD normally undergoes various revisions before it is released to the public for application in various jurisdictions. During revision, comments from various agencies and individuals are considered by FHWA through the National Committee on Uniform Traffic Control Devices (NCUTCD) in making the final rule. After reviewing and `analyzing the comments, FHWA may modify, rescind, or finalize

[1]Assistant Research Scientist, Texas A&M Transportation Institute, 701 N Post Oak Ln., # 430, Houston, TX 77024. ORCID: https://orcid.org/0000-0002-5450-1623. Email: b-kutela@tti.tamu.edu

[2]Graduate Research Assistant, Florida International Univ., 10555 W Flagler St., Miami, FL 33174 (corresponding author). ORCID: https://orcid.org/0000-0003-2697-4199. Email: hshit001@fiu.edu

[3]Assistant Professor, Texas State Univ., 601 University Dr., San Marcos, TX 78666. Email: subasish@txstate.edu

[4]Undergraduate Student, Univ. of Dodoma, Kikuyu Ave., Dodoma 41218, Tanzania. Email: lucykapaya8@gmail.com

[5]Undergraduate Student, Dept. of Statistics, Eastern Africa Statistical Training Centre, Dar es Salaam 35103, Tanzania. Email: tarimoelizaberth99@gmail.com

Note. This manuscript was submitted on November 13, 2023; approved on May 2, 2024; published online on September 12, 2024. Discussion period open until February 12, 2025; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Transportation Engineering, Part A: Systems***, © ASCE, ISSN 2473-2907.

the proposed changes (Broyhill et al. 2002) to the current version of the MUTCD. This being the case, the comments from individuals and agencies might significantly impact the MUTCD.

Focusing on the agencies, the NCUTCD has 21 sponsoring agencies, each with its own mission to accomplish (NCUTCD 2022). The comments from these sponsoring agencies may be of higher weight compared with comments from individuals due to the expertise and experience of the people involved in the review process. However, the interests and concerns of the NCUTCD sponsoring agencies have yet to be thoroughly investigated and appreciated in the existing literature. This study aims to fill this gap by identifying and understanding the unique interests of each agency as reflected in their comments. Moreover, the study explored commonalities and differences among the agencies, shedding light on the distinct agendas each agency directs within the MUTCD. The focus is on the extent to which each agency directs its agenda in the MUTCD and the level at which each agency addresses the interdisciplinary matters.

MUTCD's Preparation Process

In the 1920s, before the establishment of the MUTCD, different states used different road signs and pavement markings. Wisconsin, Minnesota, and Indiana attempted to create a basis for uniform signs and road markings. One of their pioneering works included classifying road sign shapes depending on the level of danger represented by highway situations. In 1924, AASHTO issued a report that combined all efforts to standardize signs and markings, which became the groundwork of the *Manual and Specifications for the Manufacture, Display, and Erection of U.S. Standard Road Markers and Signs*. In 1927, the *Manual on Street Traffic Signs, Signals, and Markings* was also published. However, neither manual was comprehensive, which led to the introduction of the MUTCD for streets and highways in 1935. This manual addressed signs, markings, signals, and islands (FHWA 2009).

Since then, the manual has been amended at least every decade to accommodate changes due to technology or traffic patterns. Several editions were made from that time, including in 1939, 1948, 1954, 1961, 1971, and 1978, to accommodate increased traffic, new signs, and higher speeds and improve safety. The 2009 edition is the 10th and most recent edition, with three revisions (Hawkins 2015).

Normally, the FHWA publishes a Notice of Proposed Amendments (NPA), which is usually recommended by the NCUTCD (NCUTCD 2022). This notice allows the public and federal or NCUTCD sponsoring agencies to give their opinions on proposed changes regarding MUTCD standards. NCUTCD sponsoring agencies such as AASHTO, ASCE, Institute of Transportation Engineers (ITE), American Highway Users Alliance (AHUA), National Association of City Transportation Officials (NACTO), National Association of County Engineers (NACE), International Bridge, Tunnel and Turnpike Association (IBTTA), and American Traffic Safety Services Association (ATSSA) have the opportunity to provide their comments. The 2021 NPA contained more than 600 proposed changes and received more than 25,000 comments from the public and NCUTCD sponsoring agencies (ATSSA 2021a; Paniati 2021). Comments from NCUTCD sponsoring agencies have relatively higher weights because they are likely to originate from people who utilize the MUTCD in their daily activities and there are relatively few to handle.

Sponsoring Agencies' Missions

Each agency has its own mission, which can influence the content of the MUTCD. AASHTO, responsible for all modes of transportation in all 50 states, the District of Columbia, and Puerto Rico, sets the standards for specifications, test protocols, and guidelines used in design and construction processes across the US (AASHTO 2023). ASCE is an authoritative source for codes and standards that protect the public. ASCE's mission is to protect health, wealth, and welfare while influencing the adoption of emerging technologies, analytics, and system thinking (ASCE 2022). ITE is a global membership organization of transportation professionals seeking to improve mobility and safety for all transportation system users while also assisting in developing smart and livable cities (ITE 2022). AHUA serves as the transportation community's unified voice, promoting safe, uncongested roadways and increased freedom of mobility (AHUA 2022). NACTO was founded to exchange transportation ideas, insights, and practices and to collaboratively handle national transportation challenges (NACTO 2023). NACTO also shows a high interest in public safety, especially for pedestrians. NACE is a professional nonpartisan organization that owns and manages 46% of the country's roads and 40% of its bridges, making it the single largest shareholder in local road and bridge construction, repair, expansion, and maintenance (NACE 2023). IBTTA is the global trade organization for toll facility owners and operators and the businesses that support them, founded in 1932 (IBTTA 2022). Lastly, ATSSA represents the roadway safety infrastructure business through successful legislative advocacy, traffic control safety training, and a broad member relationship (ATSSA 2021b).

The MUTCD undergoes a thorough process before publishing the final version. Such a process constitutes accommodating comments from various individuals and agencies, including the NCUTCD sponsoring agencies. However, the influence of such comments on the final version of the MUTCD is not well-explored and appreciated. This study therefore uses both supervised and unsupervised text mining analysis to understand any available patterns, differences, or commonalities in the agencies' concerns in shaping the MUTCD by using the agencies' submitted comments. This understanding can foster greater collaboration among the agencies, leading to more well-rounded and comprehensive guidelines. The study's findings also add to the existing literature on the development and implementation of traffic control standards, shedding light on the dynamics between various organizations and providing valuable insights for future revisions of the MUTCD and similar technical documents. The rest of the paper is organized as follows: the next section presents the study methodology that covers data and methods applied. Results and discussion are then presented, followed by conclusions and future studies.

Methodology

This section presents the data and methods used for analysis. It covers the data description, unsupervised text mining, and supervised text mining analytical approaches.

Data Description

More than 25,000 comments from the public and NCUTCD sponsoring agencies were collected by the FHWA through the public register by May 2021 (Paniati 2021). Even though all comments are valued for MUTCD improvement, comments from NCUTCD sponsoring agencies may be given higher consideration due to the experience with the subject matters from agencies' personnel. Thus, only comments from NCUTCD sponsoring agencies were extracted and analyzed for this study. The agencies include AASHTO, ASCE, ITE, AHUA, NACTO, NACE, IBTTA, and ATSSA. A total of 1,187 comments from NCUTCD sponsoring agencies were analyzed. Fig. 1 shows the number of comments from each NCUTCD

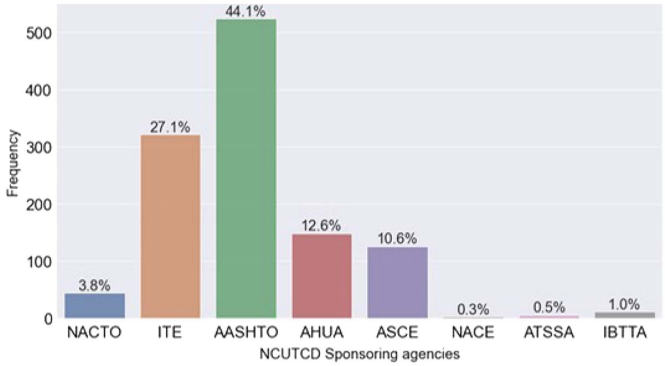


Fig. 1. Frequency distribution of NCUTCD sponsoring agencies' comments.

sponsoring agency that was analyzed. It can be observed that AASHTO had the highest number of comments (44.1%), followed by ITE (27.1%) and AHUA (12.6%). NACE and ATSAA were observed to have the lowest number of comments with 0.3% and 0.5%, respectively.

Unsupervised Text Mining

To understand the influence of the NCUTCD sponsoring agencies on the MUTCD, this study applied two text mining techniques: text network analysis (TNA) and text cluster analysis (TCA). The study used the comments (in text format) provided by personnel from various NCUTCD sponsoring agencies. TNA and TCA are among the methodologies that have been utilized to analyze text data to gain insights because they allow efficient analysis of large volumes of text data, which may require more work to analyze manually. Thus, we considered them appropriate for this study.

TNA was applied to explore the central focus of the NCUTCD sponsoring agencies by pointing out the major keywords and associated co-occurrence keywords that form a community of keywords. Fig. 2 presents the skeleton of the text network, which shows the key components of the network. TNA has been used in various fields including literature and linguistics (Hunter 2014) and traffic operations and safety (Kutela et al. 2021), among others. TNA uses nodes and edges to represent the relationship between keywords in a particular corpus of words. The size of the node represents the frequency of the keyword, and the thickness of the edges represents the co-occurrence of the keywords connected by the edge. Fig. 2 shows a typical text network with nodes, edges, and communities. Communities represent subsets of the larger network, containing closely related words distinct from other subsets in the same network. The strength of TNA lies in its ability to visualize keywords and establish a connection among them (Kutela et al. 2023; Paranyushkin 2012).

When executing TNA, several processes are performed on the data. Normalization, as the first procedure, is responsible for converting unstructured to structured data. In text normalization, any stop words, punctuation marks, and symbols are eliminated, and uppercase letters are transformed into lowercase, ensuring a consistent approach for analysis. The normalized text was then converted into a corpus of words using an algorithm under the Quanteda package version 3.0 in the R environment (Benoit et al. 2018) The outcome of this procedure is utilized to create a matrix of keywords and their corresponding occurrence frequencies. The algorithm in the Quanteda package is then used to map the keywords in a network. The keywords are presented as nodes, and the connections between keywords are shown as links. The larger the node of a keyword, the higher the frequency of its occurrence. Also, the thicker the link, the higher the co-occurrence frequency of the two connected keywords.

Although various metrics can be used for comparative analysis, this study used document frequency and collocated frequency to compare the comments from the NCUTCD sponsoring agencies. Different from the keyword frequency, which presents the number of times a keyword appears, document frequency presents the number of documents with the keyword of interest. Collocation frequency presents the number of times the keywords are located next to each other. Collocation frequency plays a major role in text clusters or community formation. A community represents keywords collectively clustered in the text network. A text network can have more than one community.

Further, TCA, which uses a simple Reinert textual clustering method, was applied (Santelli et al. 2020). Reinert's clustering method is an unsupervised approach to dividing a collection of texts into clusters based on their content. When applied to a corpus of texts, this technique can identify and extract key topics, providing insights into the underlying themes and ideas embraced by the authors (Wieser et al. 2023). The application of TCA was because the text network is incapable of clearly distending the

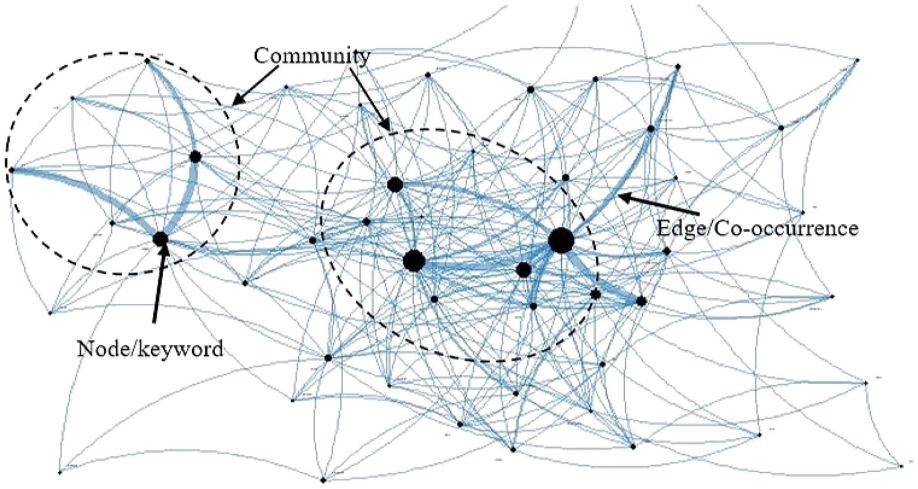


Fig. 2. Skeleton of the text network.

J. Transp. Eng., Part A: Systems, 2024, 150(11): 04024076

keywords cluster. In TCA, the keywords are arranged by the similarity of themes in each frequency, chi-square, or likelihood metric (Santelli et al. 2020). Frequency metrics were applied in this study because they provide the number of keywords in a particular cluster.

Supervised Text Mining

Both TNA and TCA are unsupervised text mining methods. The two approaches are based on the clusters of the keywords but cannot portray the keywords associated with certain agencies. To understand the key features associated with the agency, a supervised text mining method that used four text classifiers [support vector machine (SVM), random forest (RF), neural network (NN), and naïve Bayesian (NB)] was applied. The classification aims to comprehend key features affiliated with the NCUTCD sponsoring agencies' comments. The four classifiers were selected due to their ability to predict text data (Arteaga et al. 2020; Joachims 1998; Pranckevičius and Marcinkevičius 2017). The performance of these classifiers was compared based on prediction accuracy, sensitivity, and specific scores. The best-performing classifiers were used to draw conclusions on the key features that the NCUTCD sponsoring agencies presented in their comments to the MUTCD. The features were then compared across agencies to understand the agencies' influence on the revision of the MUTCD.

The comments were categorized into two categories to perform the supervised text mining. For each category, a set of comments for a given NCUTCD sponsoring agency of interest was defined. That is, if the ASCE was the NCUTCD sponsoring agency of interest, all ASCE comments were classified as one, whereas comments from other agencies were classified as zero. The same was done for AASHTO, AHUA, and ITE. The number of comments from NACTO, NACE, IBTTA, and ATSSA was relatively small, and thus, no supervised text mining was performed.

To perform the supervised text mining, it is crucial to split the data into train and test data sets. The ideal train/test split ratio is dependent on both overall data set size (n) and accuracy of classification (Muraina 2022; Dobbin and Simon 2011). For reasonably sized data sets ($n \geq 100$), with higher accuracy, the commonly employed approach of assigning two-thirds of cases to training is found to be nearly optimal (Dobbin and Simon 2011). The research team attempted three data splitting options for training and testing the supervised text mining algorithms 80%:20%, 70%:30%, and 60%:40%. This was done to determine the splitting option that does not overfit the data.

The initial analysis revealed that algorithms trained using 80% and tested using 20% of the data were susceptible to overfitting. On the other hand, the performance of the algorithm trained using 60% training set and 40% testing set resulted in poor prediction accuracy. A 70% training data set ensured robust model learning, and the 30% allocated for testing provided an adequate evaluation set, reducing the risk of overfitting and ensuring statistical significance in assessing model performance on unseen data. Thus, the data in this study was divided into two data sets: 70% for training purposes and 30% for testing purposes.

Prediction accuracy, precision, and F1 score, as indicated in Eqs. (1)–(3), respectively, were used to evaluate the performance of the four classifiers. These metrics are fundamental for assessing the performance of text mining algorithms in identifying relevant information from the data. Accuracy measures the overall correctness of our model's predictions, and precision quantifies the accuracy of the positive predictions made by the model. The F1 score, which is the harmonic mean of precision and recall, provides a balanced assessment of the model's performance in terms of both precision and recall. The score can have values between zero and one, with one being the best score (Humphrey et al. 2022). A detailed explanation of the metrics was also given by Alduayj and Rajpoot (2018) and Sokolova et al. (2006)

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \tag{1}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2}$$

$$F\text{-1 score} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \tag{3}$$

where TP = true positive, the actual positive values correctly classified as positive; TN = true negative, the actual negative values correctly classified as negative; FP = false positive, the actual negative values incorrectly classified as positive; and FN = false negative, the actual positive values incorrectly classified as negative.

Further, a set of important features for each NCUTCD sponsoring agency was established. However, only the top 20 important features from each classifier were used to interpret the results. The parameters of accuracy, precision, and F1 score will be used to identify the best classifiers that will be used to identify key phrases associated with a given agency.

Results

This section presents the results and discussions for the text analysis. The section is divided into two subsections, the unsupervised and supervised text mining results and discussions. The unsupervised text mining section covers the text network and text clusters, and the supervised text mining section covers text classifiers.

Unsupervised Text Mining Results

This section presents the results obtained from unsupervised text mining analysis to understand possible patterns or differences among NCUTCD sponsoring agents' comments. Seven text networks text clusters were produced. These text networks and clusters were for overall comments, AASHTO, AHUA, ASCE, ITE, NACTO, and a combination of NACE, IBTTA, and ATSSA.

Overall Comments

Fig. 3 presents the text network of the overall comments of the NCUTCD sponsoring agencies to the MUTCD draft. The network was centered on the keywords sign, lane, and use. Additionally, the keyword speed appeared more frequently as indicated by the size of the node. Each of these keywords is connected to several other keywords. However, there are a few key connections that can be visually observed. The keyword speed is strongly connected to limit, indicating that the comments were about the speed limit in miles per hour, as indicated by the connected keyword mph. On the other hand, the keyword sign had the biggest node, indicating the highest frequency, and was strongly linked with lane and use, implying that numerous comments focused on the use of lane signs. Other noticeable connections were between bus, lane, and marking, indicating that comments about bus lane markings were presented to a great extent. Other weak connections were observed between pedestrian and signal, traffic and control, work and sign, and pavement and markings.

Although the text network presents the relationship between keywords, the clusters to which each keyword belongs can hardly be established. Thus, Fig. 4 was created to explore the clusters of keywords for all comments. Fig. 4 presents six clusters. Of these,

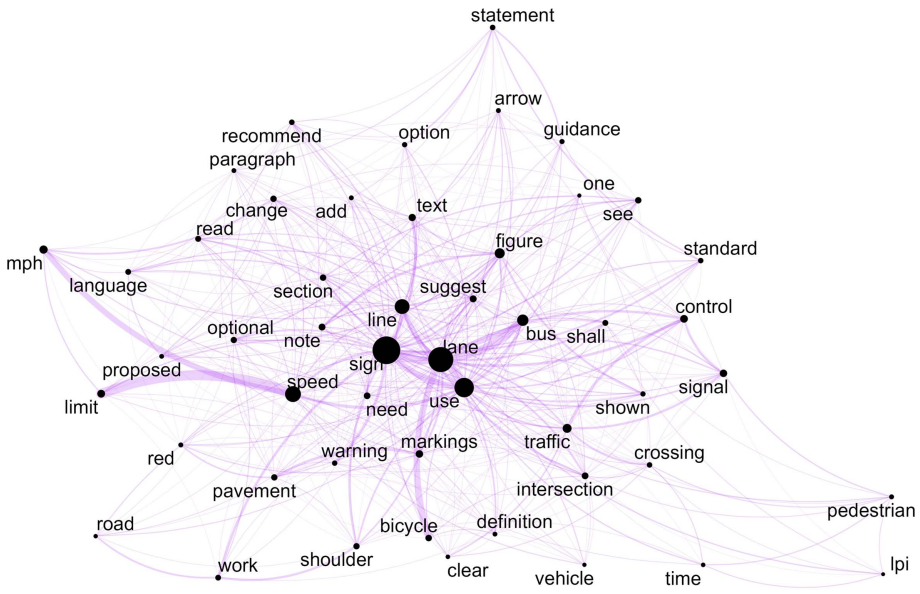


Fig. 3. Text network results for all NCUTCD sponsoring agencies' comments.

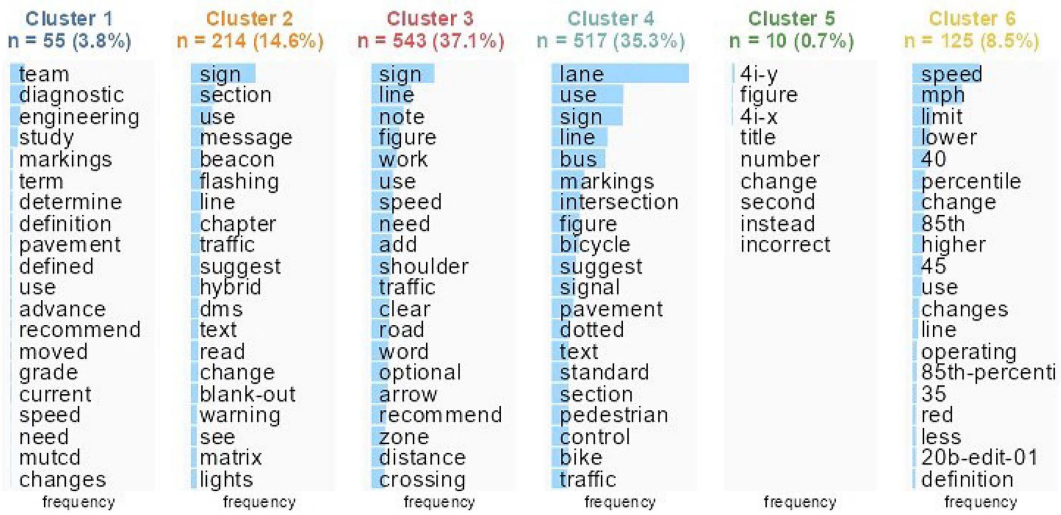


Fig. 4. Text clusters results for all NCUTCD sponsoring agencies' comments.

more observations/comments were in Cluster 3 (37.1%), followed by Cluster 4 (35.3%) and Cluster 2 (14.6%). Cluster 3 was observed to focus on the signs and signals on lanes and shoulders. The same cluster presented comments on pedestrians and crossing. Cluster 4, on the other hand, pointed out the lane use markings and signs for buses and bicycles. Such pavement marking can be at the intersections. Cluster 2 focused on the warning message signs in terms of beacons, dynamic message signs, and other lights.

The remaining clusters presented a relatively small number of comments, but contained relatively useful information. Cluster 6 was all about speed limits, whereby lower, higher, operating, and 85th percentile speed limit dictate the cluster. It is vivid that 40 and 45 mi/h speed limits have extensively been discussed in the comments. Cluster 1 covered the diagnostic and engineering studies necessary for pavement and markings. In the same cluster, various recommendations are presented to change the current document to the proposed content of the MUTCD. Lastly, Cluster 5 presented the changes proposed for figures, tables, and titles.

The text networks and associated clusters do provide the facts for the overall comments from all NCUTCD sponsoring agencies. However, to understand the key theme presented by each agency, the comments ought to be subdivided according to the NCUTCD sponsoring agency.

AASHTO-Based Comments

Fig. 5 presents the text network for AASHTO comments. The network was centered on the keyword sign, with three other keywords, namely, use, lane, and speed, showing higher frequency as indicated by the node size. The connection between keywords provides more information of the key focus of the AASHTO-based comments.

The co-occurrence of speed-limit and speed-mph indicates that a significant number of comments focused on the speed limit. For instance, one of the actual comments from AASHTO stated that "Recommend 'may' instead of 'should.' If speeds are higher than the speed limit, agencies should take measures to reduce the speeds

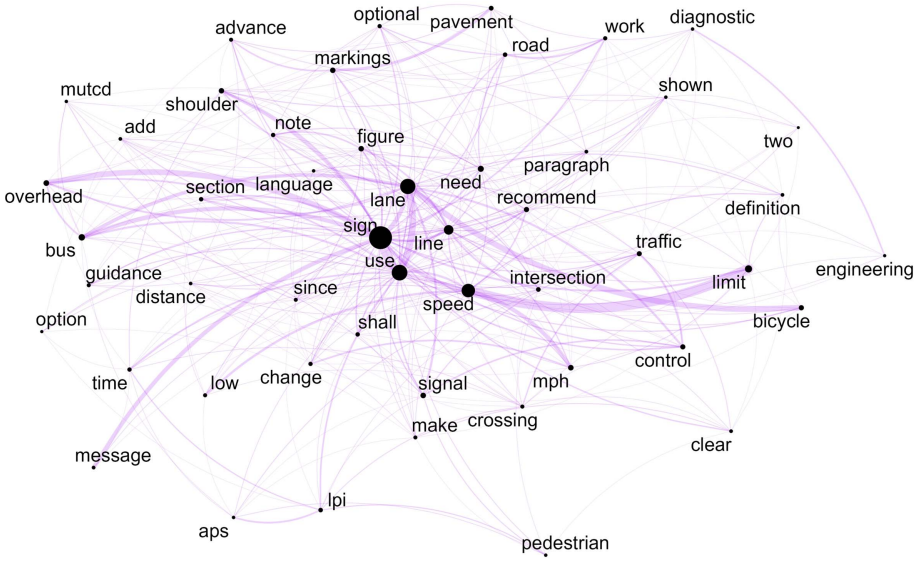


Fig. 5. Text network from AASHTO comments.

or raise the speed limit. We shouldn't be making it easier to continue excessive speeding." Another AASHTO comment related to speed limits stated, "Disagree with option #2 for a VSL sign. While the numerals could be positive or negative contrast depending on the CMS technology (black or white background), the words 'Speed Limit' should always be black-on-white." This implies a greater concern about speed limits from the AASHTO agency.

The co-occurrence of overhead, message, and sign indicates the presence of a substantial number of comments that showed concern about the overhead message signs. Further, the co-occurrences of bus-sign, use-lane, and bicycle-lane are among the co-occurrences that are dominant in the network. Additionally, the network presents keywords such as pedestrians, signal, pavement, and markings, which are necessary for safety and operations. The network also has some keywords that are used for editorial purposes of the MUTCD, such as paragraph, language, figure, note, and so on. Lastly, guidance and optional keywords, which are major parts of the MUTCD, are included in the network. The text network shows the concern of AASHTO over some modes of transportation, with keywords including sign, bus, and bicycle.

AHUA-Based Comments

Fig. 6 presents the text network for AHUA comments. Like AASHTO, the AHUA network was centered on signs and has lane as a major keyword. The AHUA text network, however, had the strongest links for bus-lane, traffic-control, message-sign, flash-beacon, work-use, and low-speed. The presence of such keywords within the AHUA comments indicates a significant concern over highway/roadway use and safety, which underscores the organization's commitment to promoting safe highways and enhanced freedom of mobility (AHUA 2022). The network also presented keywords related to modifying the MUTCD draft.

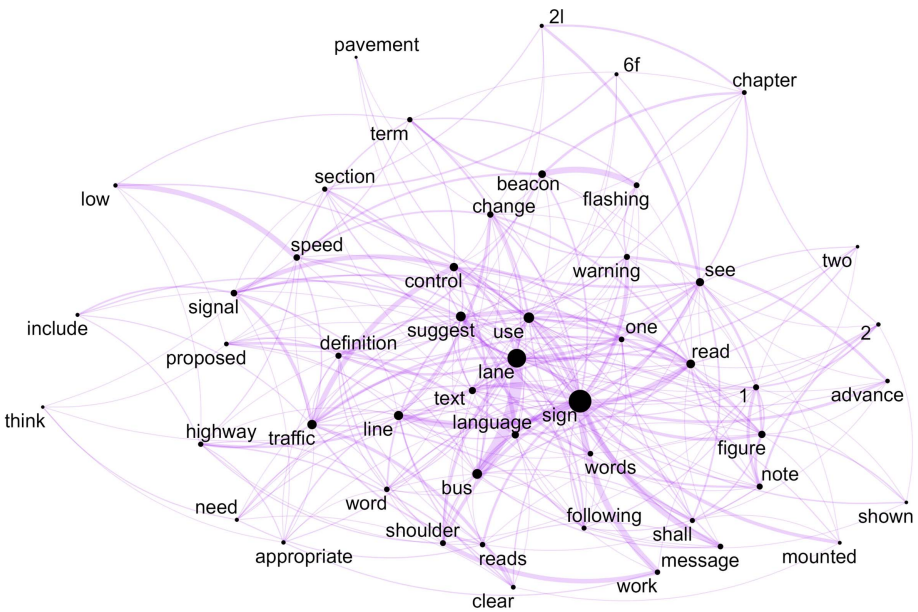


Fig. 6. Text network from AHUA comments.

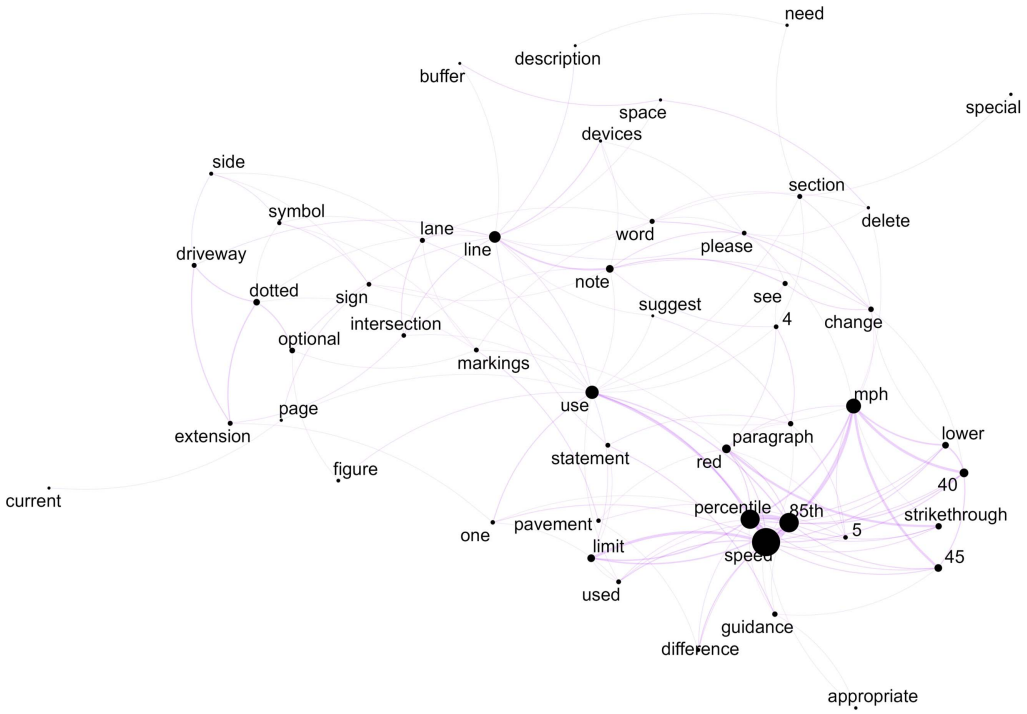


Fig. 7. Text network from ASCE comments.

Such keywords include suggest, language, propose, and appropriate, among others.

ASCE-Based Comments

Fig. 7 presents the ASCE-based comments. According to the text network, speed limit–related terms were the center of discussion for the ASCE comments. These included keywords such as speed, percentile, 85th, mph, and lower. The 85th percentile and speed keywords were observed to be the center of discussion. Apart from that, a few scattered roadway-related keywords were also observed, including pavements, marking, driveways, and intersections, among others.

Contrary to AASHTO and AHUA comments, ASCE comments did not focus much on signs, lane, and use. These keywords had a relatively small node size indicating that they appear less frequently. The ASCE network, however, contained a significant number of keywords that focused on changing/editing the current MUTCD draft. Such keywords included strike through, add, change, statement, description, delete, and please.

ITE-Based Comments

Fig. 8 shows the text network for ITE comments. The figure is centered on the three keywords sign, lane, and line, similar to AASHTO and AHUA. Furthermore, speed limit–related keywords

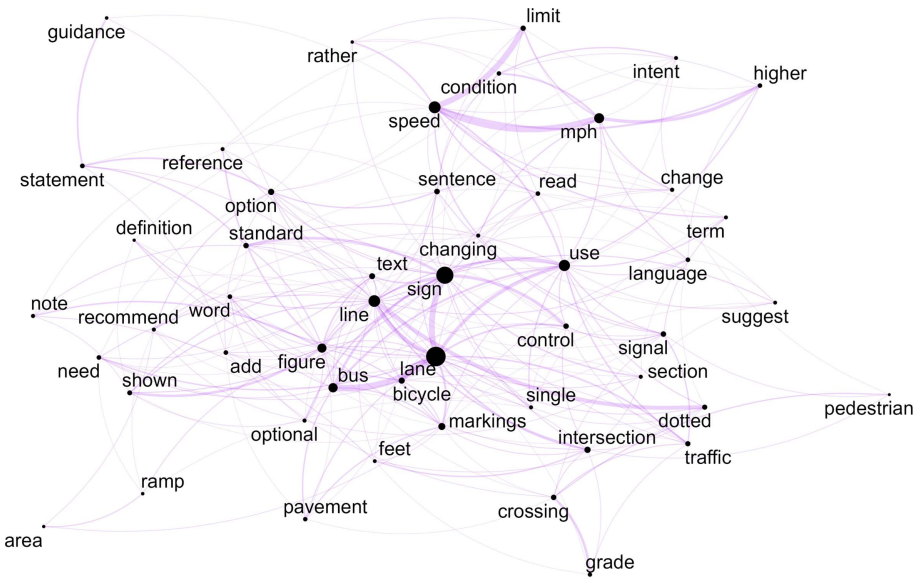


Fig. 8. Text network from ITE comments.

J. Transp. Eng., Part A: Systems, 2024, 150(11): 04024076

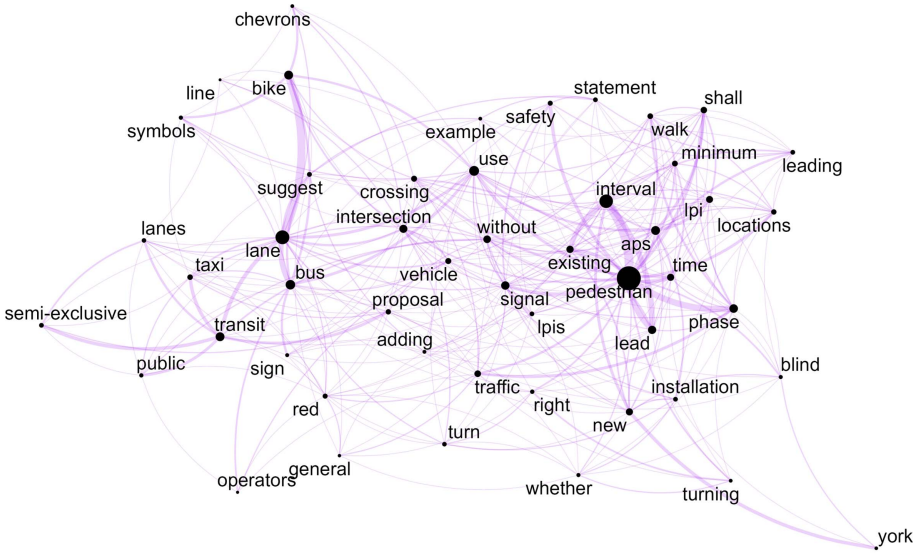


Fig. 9. Text network from NACTO comments.

were dominant in this network, similar to ASCE and AASHTO. The additional scattered keywords, such as intersection, signals, markings, and the keywords for modifying the MUTCD draft, were similar to the ones in the previous networks.

NACTO-Based Comments

Fig. 9 presents the text network for NACTO-based comments and shows a significant difference from the previous networks. The main keyword in the network was pedestrian, and the major co-occurring keywords were pedestrians, intersection, interval, lead, and bike. This observation signifies that the NACTO-based comments focused on pedestrians and bicyclists. In fact, several keywords point out that NACTO comments insisted on pedestrian and bicyclist needs. The presence of keywords such as walk, LPI (leading pedestrian interval), leading, and blind, among others, indicated the design parameters for pedestrian and bicyclist facilities. In addition, the network also presented transit-related keywords such as exclusive transit lanes. The keywords statement, example, adding, and shall were among the keywords that indicate the directions to the edits of the draft.

NACE, IBTTA, and ATSSA-Based Comments

Lastly, the text network in Fig. 10 presents the keywords for other NCUTCD sponsoring agencies which include NACE, IBTTA, and ATSSA. Overall, the text network showed a relatively small number of comments to be individually represented. The network was centered on the keywords arrow, text, board, figure, and optional.

The keywords are indicative of the desire to edit the MUTCD draft. However, the keywords do not provide a clear understanding of key intentions. The other community on the top left presents the keywords that show the requirements for decisions, which should be based on engineering diagnostic studies.

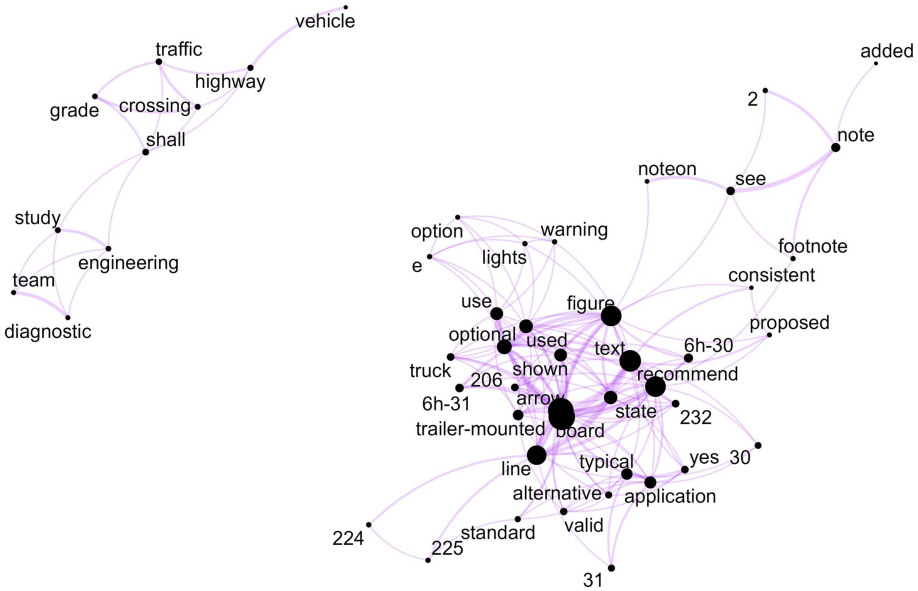


Fig. 10. Text network from other agency (NACE, IBTTA, and ATSSA) comments.

J. Transp. Eng., Part A: Systems, 2024, 150(11): 04024076

The information presented in the text networks and clusters aims to understand the patterns of the keywords for the NCUTCD sponsoring agencies' comments. To understand the key features that are specifically associated with a given NCUTCD sponsoring agency, four supervised text mining techniques were analyzed and discussed as shown in the following section.

Supervised Text Mining Results

This section presents the results and discussion of the supervised text mining analysis using text classifiers. Table 1 presents the performance metrics of different classifiers for the given agencies. It can be observed that the prediction accuracies of naïve Bayes were relatively lower than the prediction accuracies for all other algorithms. However, this analysis focuses on identifying the key phrases associated with a given agency. Thus, precision is a more prominent metric. According to the results in Table 1, RF, SVM, and neural network (NNet) classifiers consistently produced higher precision scores irrespective of the originality of the comments. Thus, the important features from these three classifiers were used to draw conclusions about the key features of the agencies' comments.

Fig. 11 presents the 20 topmost keywords from four NCUTCD sponsoring agencies. Fig. 11(a) shows the topmost important features for comments from AASHTO, and Fig. 11(b) shows those from AHUA. According to the results, several keywords are common in each NCUTCD sponsoring agency. Considering all classifiers for the four agencies, the keyword sign appears highly important among all agencies. This implies that road signs are a critical issue observed by all agencies and that a change might be desired on the existing road signs. Keywords such as 85 or percentile and shoulder appear in most agencies' comments, except for AHUA in Fig. 11(b). The implication of this is that irrespective of the focus of the NCUTCD sponsoring agency, the 85th percentile speed and the roadway shoulder operations appear to be a cross-cutting issue that is addressed by all the agencies. Other common keywords observed include recommend, suggest, and shoulder. The presence of common recommendations and suggestions implies that the NCUTCD sponsoring agencies have a common interest in looking forward to changes from suggestions/comments given by the FHWA for MUTCD improvement.

The RF classifier was observed to have a relatively higher precision and accuracy score compared with the other classifiers and, hence, was used to identify peculiar concerns among the agencies.

Table 1. Performance metrics of the text classifiers

Agency	Performance metrics	Classifier SVM (%)	RF (%)	NNet (%)	NB (%)
AHUA	Accuracy	84.9	85.2	81.8	23.6
	Precision	44.4	44.4	41.3	14.9
	F1 score	23.2	13.3	49.2	25.6
AASHTO	Accuracy	59.0	65.5	60.1	53.3
	Precision	56.4	69.1	57.1	49.7
	F1 score	53.8	55.7	57.1	52.3
ASCE	Accuracy	90.9	92.6	87.5	11.4
	Precision	40.0	70.0	33.3	8.8
	F1 score	20.0	35.0	38.9	16.2
ITE	Accuracy	70.1	72.6	62.4	46.2
	Precision	34.1	44.1	30.8	30.4
	F1 score	21.1	23.8	32.7	100.0

For instance, ASCE, as shown in Fig. 11(c), is more represented by speed-based texts such as 85th percentile, speed, percentile, and percentile speed. The implication is that the ASCE comments suggest several modifications to the definition of the 85th percentile speed in the MUTCD. For instance, an ASCE comment stated, "'40 mph or lower' works here because they are discrete speeds, NO 85th Percentile Speed." Moreover, another comment stated "The introduction of 85th percentile speed here seems to provide an inconsistency with the speed specified in the 2 equations in paragraph 14. The difference between speed limit and 85th percentile speed could point to using different equations." The presence of such comments shows a greater concern by the ASCE on the use and definition of the 85th percentile.

Changeable message signs (CMS), lanes, buses, intersections, and traffic appeared to be important concerns for the AASHTO agency. This implies an interest in day-to-day traffic operations and the use of intelligent transportation systems (ITS). One of the comments from the AASHTO agency stated, "This language seems to indicate that line matrix signs are now ONLY to be used for temporary traffic control. Consider that many existing permanent changeable message signs are line matrix signs and are used for purposes other than temporary traffic control." Another comment related to bus operations stated, "Consider enhancing bus markings within gore areas due to the number of potential conflicts with buses in shoulder." These comments imply a greater concern of AASHTO on the use of line matrix signs and bus operations.

Lane was also observed to be an important keyword under the AHUA agency, which shows a similar interest to the AASHTO agency. Some of the comments showing concern for lanes stated, "What is a 'semi-exclusive bus lane'?" and another stated, "Although 'channelizing devices' is not a defined term, Section 6F.63 makes it clear that drums, tubular markers, vertical panels, cones, and Type 1, 2, and 3 barricades are all channelizing devices. A novice MUTCD user who reads this proposed definition could easily think that all the above types of channelizing devices are appropriate for creating a separate bicycle lane. I doubt that is what the Bicycle Committee intends." The comments reflect the importance or query for clearance on buses, bicycles, managed, and other types of lanes shown by the AHUA agency.

Guidance statement, roadway, and definition appeared to be important keywords for the ASCE agency. Like AASHTO and AHUA, lanes appeared to be a concern for the ITE. Fig. 11(d) shows the important features for comments from ITE. Dynamic message signs (DMS), also known as CMS, also appeared under the ITE agency. One of the comments from the ITE regarding dynamic message signs stated.

I suggest "All dynamic message sign, hybrid sign and blank-out sign shall be subject to meet the design requirements of paragraphs 06 through 09 in terms of shape, color, size and font. They shall display an exact duplicate of the standard static sign or other sign legends use standard symbols, the Standard Alphabets and letter forms, route shields, and other typical sign legends elements with no apparent loss of resolution or recognition to the road user when compared with static versions of the same sign or legends, except as noted in Section 2L.04 for hybrid and blank-out sign.

This implies the need for DMS to be consistent with static signs to reduce confusion or uncertainty among motorists. Other important keywords for the ITE agency included standard, ramp, lights, guidance, and option. From the supervised mining text results, it can be inferred that lanes are a common concern among AASHTO, AHUA, and ITE, and the use of dynamic or changeable message signs is a common concern between AASHTO and ITE.

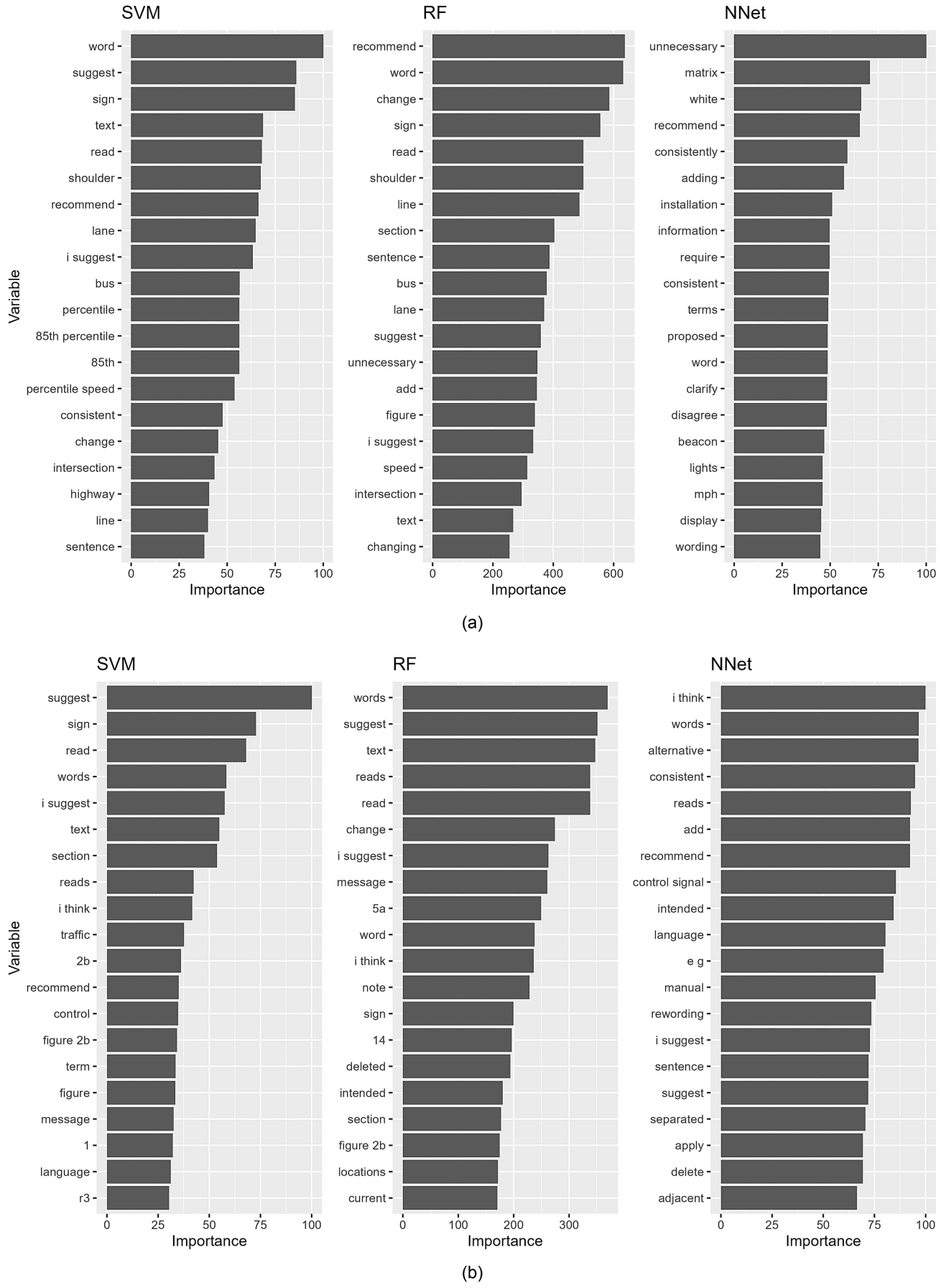
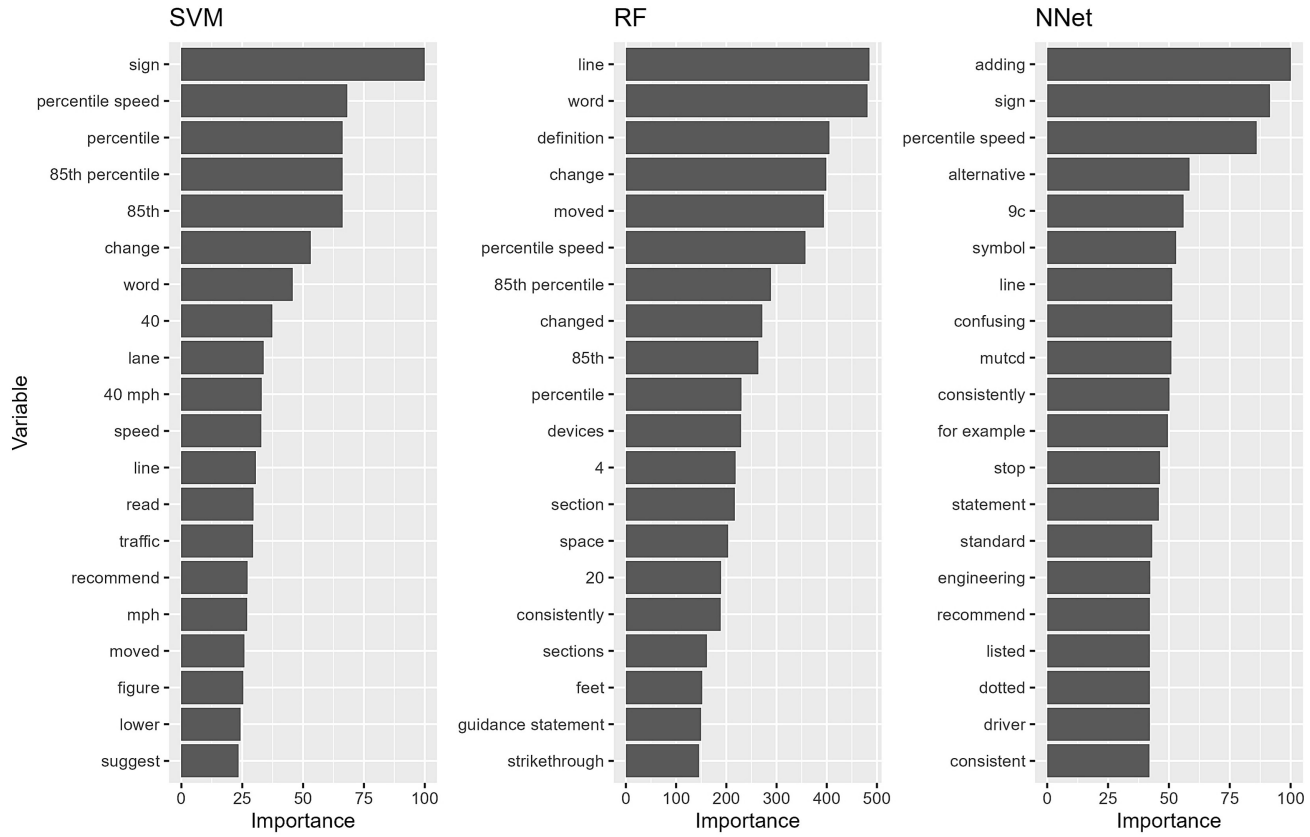
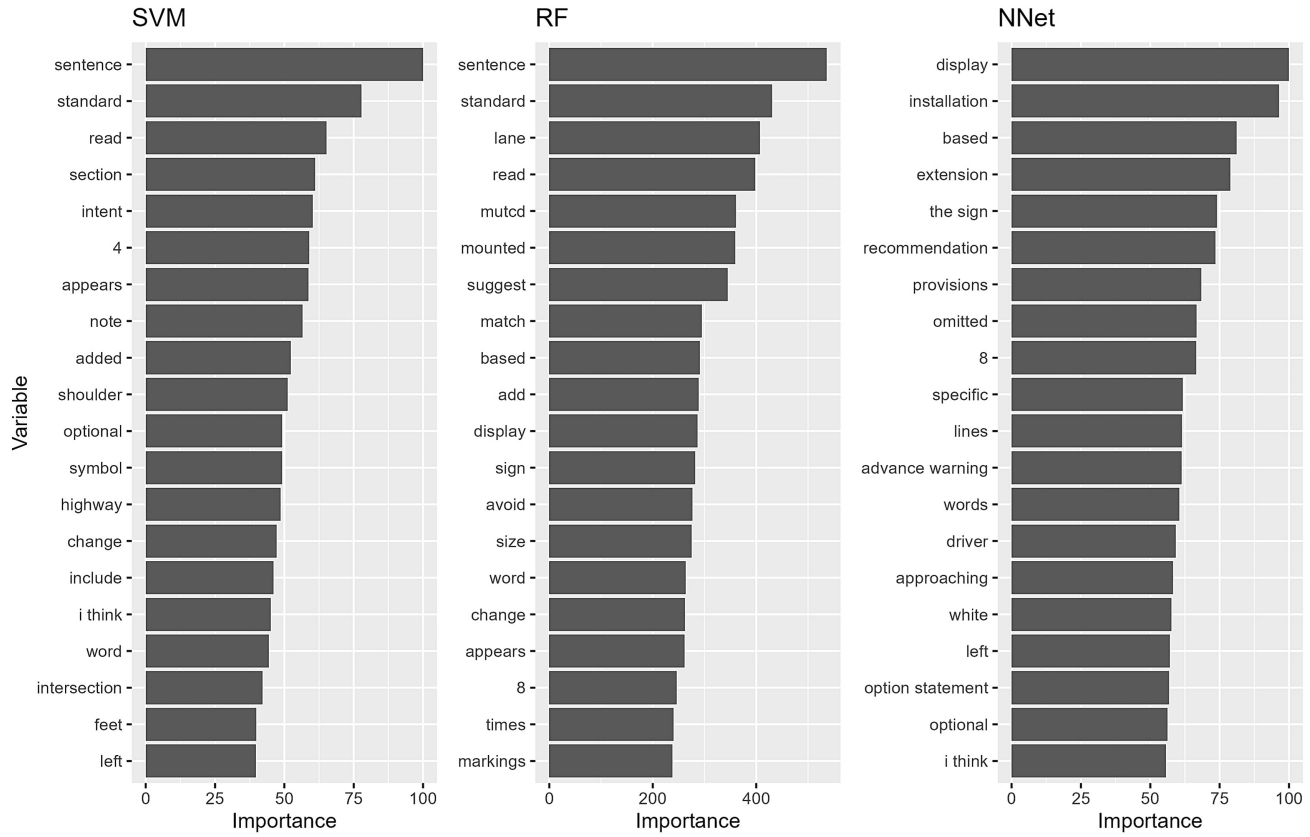


Fig. 11. Top 20 important features for comments from: (a) AASHTO; (b) AHUA; (c) ASCE; and (d) ITE.



(c)



(d)

Fig. 11. (*Continued.*)

Conclusions

This study presents a detailed analysis of the draft MUTCD comments from NCUTCD sponsoring agencies. The objective was to evaluate the presence of the agencies' interests in the MUTCD draft comments by analyzing existing patterns in similarity or differences. The study applied unsupervised and multiple supervised text mining on the draft comments from the NCUTCD sponsoring agencies to the FHWA on MUTCD comments. The NCUTCD sponsoring agencies' comments analyzed include those from AASHTO, ASCE, ITE, AHUA, NACTO, NACE, IBTTA, and ATSSA. The unsupervised text mining applied were text network and text cluster, whereas supervised text classifiers applied were RF, NB, NN, and SVM. RF, NN, and SVM were observed to perform better than other classifiers and were therefore applied to conclude on the key features of the agencies' comments.

Overall, the results indicated the presence of various shared interests across NCUTCD sponsoring agencies and some unique features that were specific for the agencies. The text network results revealed that comments related to speed limit, signs, bus, and bicycle lanes were dominant for AASHTO, ASCE, and ITE. NACTO comments focused more on pedestrians, bicycles, and transit-related items such as lead pedestrians' signs, bike lane, and semi-exclusive transit signs. AHUA, on the other hand, focused on bus lane signs and traffic control.

The supervised text mining results indicated a similar trend using the RF classifier, which had a relatively higher accuracy and precision score compared with other classifiers. Signs, 85th percentile speed, and shoulder were observed to be dominant among AASHTO, ASCE, AHUA, and ITE. On the other hand, unique patterns were also observed where AHUA comments showed much concern about the modification of the MUTCD. AASHTO comments showed a particular concern on the standards and use of changeable message signs, lanes, traffic, and intersections, whereas those from AHUA showed an interest in the definition and use of different types of lanes, including bicycle, managed, and bus lanes. ITE comments showed particular concern about lanes, dynamic message signs, ramps, lights, and guidance. On the other hand, ASCE showed peculiar concerns about roadways and definitions.

The findings from this study can be used to understand the major concerns of different NCUTCD sponsoring agencies regarding the revision of the current MUTCD. Different agencies were observed to have different concerns beyond their organization's visions. However, it is clearly observed that the majority of the agencies have a great concern for the safety and well-being of society by showing a common interest in safety influencing factors such as speed, road signs, and the definition and use of the 85th percentile speed. Moreover, AASHTO and ITE had a similar concern regarding the use of DMS, whereas AASHTO and AHUA had a similar concern regarding lanes. Further study can be done to analyze the patterns of similarity or differences among the comments from the public (excluding the NCUTCD sponsoring agencies) in relation to geographic location, occupancy, sociodemographic factors, or other criteria.

Study Limitations and Future Studies

In conducting this study, several limitations need to be acknowledged to provide a comprehensive understanding of the study's scope and potential constraints. Firstly, the study did not identify comments from sponsoring agencies that were incorporated into the manual. Rather, the study explored the comments given by the agencies to determine if they align with the agencies' missions, visions, or general societal needs.

Furthermore, the study relied solely on recent comments obtained by the FHWA regarding the MUTCD improvement as of May 2021. This limited time frame for data collection may have introduced biases into the conclusions drawn. Therefore, it is imperative to conduct additional research spanning different time periods to accurately discern any shifts in agencies' concerns over time. This approach would offer a more comprehensive understanding of evolving perspectives within the field, enhancing the validity and relevance of subsequent findings.

It is essential to recognize that despite these limitations, our study contributes valuable insights regarding the MUTCD revision. To address these constraints, future research endeavors could explore sponsoring agencies' comments over multiple revisions providing more understanding of how interests may have shifted or remained consistent over time.

Data Availability Statement

Some or all data, models, or codes that support the findings of this study are available from the corresponding author upon reasonable request.

References

AASHTO. 2023. "American association of state highway and transportation officials." Accessed June 10, 2023. https://www.transportation.org/home/organization/.

AHUA (American Highway Users Alliance). 2022. "American highway users alliance." Accessed March 21, 2022. https://www.highways.org/.

Alduayj, S. S., and K. Rajpoot. 2018. "Predicting employee attrition using machine learning." In *Proc., 2018 13th Int. Conf. on Innovations in Information Technology, IIT 2018*, 93–98. New York: IEEE. https://doi.org/10.1109/INNOVATIONS.2018.8605976.

Arteaga, C., A. Paz, and J. W. Park. 2020. "Injury severity on traffic crashes: A text mining with an interpretable machine-learning approach." *Saf. Sci.* 132 (Dec): 104988. https://doi.org/10.1016/j.ssci.2020.104988.

ASCE. 2022. "American Society of Civil Engineers." Accessed April 13, 2022. https://www.asce.org/.

ATSSA (American Traffic Safety Services Association). 2021a. "ATTSA submits 120 comments for feedback to proposed MUTCD." Accessed August 27, 2021. https://atssadev.atssa.com/Technical-Services/Temporary-Traffic-Control/atssa-submits-120-comments-for-feedback-to-proposed-mutcd.

ATSSA (American Traffic Safety Services Association). 2021b. "ATTSA submits 120 comments for feedback to proposed MUTCD." Accessed February 20, 2023. https://atssadev.atssa.com/.

Benoit, K., K. Watanabe, H. Wang, P. Nulty, A. Obeng, S. Müller, and A. Matsuo. 2018. "Quanteda: An R package for the quantitative analysis of textual data." *J. Open Source Software* 3 (30): 774. https://doi.org/10.21105/joss.00774.

Broyhill, T., C. Tan Esse, and L. Ward. 2002. "Innovative traffic control devices—The rulemaking process and public comments (Part II)." *ITE J.* 72 (2): 24–26.

Dobbin, K., and R. Simon. 2011. "Optimally splitting cases for training and testing high dimensional classifiers." *BMC Med. Genomics* 4 (Apr): 1–8. https://doi.org/10.1186/1755-8794-4-31.

FHWA (Federal Highway Administration). 2009. *Manual on uniform traffic control devices (MUTCD)*. Washington, DC: FHWA.

Hawkins, G. 2015. "The MUTCD turns 80: Time for a makeover?" *Inst. Transp. Eng. J.* 85 (11): 14–16.

Humphrey, A., W. Kuberski, J. Bialek, N. Perrakis, W. Cools, N. Nuyttens, H. Elakhrass, and P. A. C. Cunha. 2022. "Machine-learning classification of astronomical sources: Estimating F1-score in the absence of ground truth." *Mon. Not. R. Astron. Soc. Lett.* 517 (1): L116–L120. https://doi.org/10.1093/mnrasl/slac120.

Hunter, S. 2014. "A novel method of network text analysis." *Open J. Mod. Ling.* 4 (02): 350–366. https://doi.org/10.4236/ojml.2014.42028.
IBTTA (International Bridge, Tunnel and Turnpike Association). 2022. "International Bridge, Tunnel & Turnpike Association." Accessed June 1, 2022. https://www.ibtta.org/about-us.
ITE (Institute of Transportation Engineers). 2022. "Institute of Transportation Engineers." Accessed June 1, 2022. https://www.ite.org/about-ite/about-ite/.
Jewell, S. 2017. *Future of the MUTCD.* West Lafayette, IN: Purdue Univ.
Joachims, T. 1998. "Text categorization with support vector machines: Learning with many relevant features." In *Proc., European Conf. on Machine Learning*, 137–142. Berlin: Springer. https://doi.org/10.1007/bfb0026683.
Kutela, B., C. Kadeha, R. T. Magehema, R. E. Avelar, and P. Alluri. 2023. "Leveraging text mining approach to explore research roadmap and future direction of wrong-way driving crash studies." *Data Inf. Manage.* 8 (1): 100044. https://doi.org/10.1016/j.dim.2023.100044.
Kutela, B., N. Novat, and N. Langa. 2021. "Exploring geographical distribution of transportation research themes related to COVID-19 using text network approach." *Sustainable Cities Soc.* 67 (Apr): 102729. https://doi.org/10.1016/j.scs.2021.102729.
Muraina, I. 2022. "Ideal dataset splitting ratios in machine learning algorithms: General concerns for data scientists and data analysts." In *Proc., 7th Int. Mardin Artuklu Scientific Research Conf.* Ankara, Turkey: Institute of Economic Development and Social Research.
NACE (National Association of County Engineers). 2023. "National Association of County Engineers." Accessed April 5, 2024. https://www.countyengineers.org/.
NACTO (National Association of City Transportation Officials). 2023. "National Association of City Transportation Officials." Accessed June 1, 2022. https://nacto.org/.
NCUTCD (National Committee on Uniform Traffic Control Devices). 2022. "Public and professional awareness of the principles of safe traffic control devices and practices." Accessed May 1, 2022. https://ncutcd.org/.
Paniati, J. F. 2021. "Setting a new course for the MUTCD." *J. Inst. Transp. Eng.* 91 (7): 6.
Paranyushkin, D. 2012. *Visualization of text's polysingularity using network analysis*. Berlin: Nodus Labs.
Pranckevičius, T., and V. Marcinkevičius. 2017. "Comparison of naïve Bayes, random forest, decision tree, support vector machines, and logistic regression classifiers for text reviews classification." *Baltic J. Modern Comput.* 5 (2): 221–232. https://doi.org/10.22364/bjmc.2017.5.2.05.
Santelli, F., G. Ragozini, and M. Musella. 2020. "What volunteers do? A textual analysis of voluntary activities in the Italian context." In *Studies in classification, data analysis, and knowledge organization*, edited by D. F. Iezzi, D. Mayaffre, and M. Misuraca, 265–276. Cham, Switzerland: Springer Nature. https://doi.org/10.1007/978-3-030-52680-1_21.
Sokolova, M., N. Japkowicz, and S. Szpakowicz. 2006. "Beyond accuracy, F-score and ROC: A family of discriminant measures for performance evaluation." In *Proc., Australasian Joint Conf. on Artificial Intelligence*, 1015–1021. Berlin: Springer.
Wieser, M., K. Schöffmann, D. Stefanics, A. Bollin, and S. Pasterk. 2023. "Investigating the role of ChatGPT in supporting text-based programming education for students and teachers." In *Proc., Int. Conf. on Informatics in Schools: Situation, Evolution, and Perspectives*. Cham, Switzerland: Springer Nature. https://doi.org/10.1007/978-3-031-44900-0_4.