TRR
JOURNAL OF THE TRANSPORTATION RESEARCH BOARD

Research Article

Transportation Research Record
1–10

DOI: 10.1177/0361198120987230
journals.sagepub.com/home/trr
SAGE

Exploratory Analysis of Unmanned Aircraft Sightings using Text Mining

Subasish Das[1]

Abstract

Because of recent technological advancements, a growing number of unmanned aircraft systems (UASs) are anticipated to occupy the U.S. National Airspace System (NAS) and operate side-by-side with human pilot controlled civil aircraft. UAS technology has transitioned to broader applications, including commercial, scientific, and expanded military use. There have been significant challenges concerning the safe and suitable integration of UASs with existing systems. The interaction between humans and increasingly automated systems is of concern to researchers. Additionally, the number of UAS sightings has increased significantly during the last few years. In this study, the research team compiled 7,400 reports of UAS sightings (2015–2018). The Latent Dirichlet Allocation (LDA) method was then applied to develop topics relevant to UAS sighting incidents. This study also developed an online interactive tool to show keywords associated with different topics. These interactive topic models can help policymakers establish new policies and regulations to address specific safety concerns.

In the past 30 years, the most significant entrant by far to the U.S. National Airspace System (NAS) has been unmanned aircraft systems (UASs). New technological advancements have accounted for negative human factors of remote aircraft operation, such as delayed information transferal or sensory information loss. To mitigate any possible safety issues and encourage efficient use of UASs in and around U.S. airports, the Federal Aviation Administration (FAA) has implemented several new policies. In addition to the FAA's rules, airports must individually develop performance measures to evaluate their capabilities and strategies to effectively incorporate UASs in their daily operations.

UASs can be used for data collection (e.g., turn movement count at an intersection), infrastructure inspections and assessments (e.g., bridge inspection), security enhancement, and autonomous delivery systems. Another essential function of UASs is that they can be used to improve emergency response systems by speeding up rescue and recovery operations for natural disasters. The recent increase of active UASs is primarily because of an increase in the number of UAS hobbyists and commercial multi-UAS frameworks. Additionally, small unmanned aircraft system (sUAS) operations are rapidly growing, and civilian use is becoming more widespread. According to the FAA estimations, there will be around 2.75 million to 4.47 million sUAS in U.S. airspace by 2021 (*1*). As airspace becomes increasingly limited, particularly in dense urban areas, the development and implementation of modern collision avoidance techniques are crucial to monitor and manage UASs. The responsibility partially falls on the FAA to prepare for this increase in UAS operations. If the FAA does not prepare the airspace for safe operation in the airfield, there will be costly delays that will take away potential benefits from stakeholders. An important aspect of implementing these new technologies is monitoring the potential impacts, beneficial or otherwise, on airport operations with standardized performance metrics. Additionally, this allows airports that have already integrated UASs into their daily operations to provide insight into the challenges and opportunities that they present.

The present study encompasses a comprehensive analysis of UAS sightings. The data presented in this study includes 7,400 incidents involving UASs in the U.S. NAS from January 1, 2015, to December 31, 2018. It includes key information about each UAS sighting event, including the city, state, time, date, and narrative report of the event. After performing exploratory data analysis, the research team applied Latent Dirichlet Allocation (LDA) to determine the key topics for each year. This paper is organized as follows. A brief review of relevant studies is

[1]Texas A&M Transportation Institute, Bryan, TX

Corresponding Author:
Subasish Das, s-das@tti.tamu.edu

described in the next section. It is followed by a discussion of the methodology and the modeling results. At the end of this paper, conclusions and recommendations are included.

Literature Review

As UASs have become more prevalent in recent years, there has been an increase in research on policy issues related to UAS technology implementation.

Airspace has become increasingly limited as UAS flight has become more prevalent, so modern collision avoidance techniques have become increasingly important. The state-of-the-art studies cover a variety of problems and solutions: a cloud-based web application to monitor real-time UAS management; a survey to assess participants' knowledge, attitude, and practices concerning UASs; an examination of human factors of UASs; control interfaces of UASs; identification of threats from UASs; safe control of UASs; a model to assess the risks associated with UASs integrated into the NAS; investigation of flying habits of UAS hobbyists; estimation of carbon dioxide (CO^2) emissions and vehicle-miles traveled (VMT) measures for UASs; and quantitative evaluation of air traffic management methods based on safety, efficiency, and computational intensity (*2–12*).

Gettinger and Michel conducted a comprehensive and detailed study in which they analyzed incidents involving unmanned and manned aircraft in the U.S. NAS (*13*). The study collected data for 921 incidents from 2013 to 2015. Two major groups were considered for the analysis: (1) sightings cases in which a pilot or an air traffic controller noticed a UAS flying within or near to their flight path, and (2) close encounters where a manned aircraft came close enough to a UAS that it met the FAA's definition of a "near mid-air collision." This study reported 65% and 35% of sightings and close encounters, respectively.

The findings show that incidents mainly occur at locations with high manned air traffic and where UAS use is prohibited.

The U.S. Government Accountability Office (GAO) created a report for Congress to address the safety concerns of sUAS use and describe the FAA's and other agencies' efforts to address these concerns (*14*). This report examined the availability of supporting data and potential safety risks associated with UASs. The report suggested that the FAA should establish a method to certify that the FAA's management of safety hazards posed by UASs that follows all possible options to improve safety.

The literature review reveals that text mining on UAS sighting reports has not been conducted yet. This study aims to mitigate this critical gap.

Methodology

Data Collection

During the last few years, reporting on UAS sightings has significantly increased. The FAA now acquires approximately 100 reports per month. One of the straightforward messages that the FAA tries to circulate is that operating UASs (without following FAA regulations) around airplanes, helicopters, and airports is dangerous and illegal. Any unauthorized activity with UASs (not following FAA regulations) is subject to criminal charges and associated fines. This study collected reports of UAS sightings from 2015 to 2018. A total of 7,400 sightings have been reported in these 4 years. Table 1 lists samples of three UAS sighting reporting information. The final dataset is publicly available at a GitHub repo (https://github.com/subasish/UASSightings).

Descriptive Statistics

Figure 1 displays the amount of UAS sightings in each state per year from 2015 to 2018. The states with a high number of UAS sightings are California (1,405 sightings), Florida (846 sightings), New York (731 sightings), and Texas (574 sightings). Other than these four states, all other states had less than 100 sightings each year. The U.S. territories that are not shown in the figure are listed below in the description.

Figure 2 displays the number of UAS sightings for each month from 2015 to 2018. It shows an overall increasing trend from year to year, as 2015 has the least total sightings and 2018 has the most sightings. Another trend shown in the figure is that the greatest number of sightings typically take place during the summer months, as well as during late spring and early fall (April to October). This trend is likely because of nicer weather that encourages people to go outside and operate UASs. The maximum number of sightings shown on the table is 333 for June 2018; the minimum number shown is 26 sightings for January 2015.

Figure 3 displays the number of sightings for the days of the week each month. As shown in Figure 2, it also illustrates a greater number of sightings in the summer months. Additionally, Figure 3 shows a significantly greater number of sightings on the weekend (i.e., Saturday and Sunday) than on weekdays (i.e., Monday to Friday). The maximum shown in the figure is 219 sightings on Sundays in August, while the minimum number is 28 sightings on Tuesdays in January.

Figure 4 shows the total number of UAS sightings for each day of the week over 4 years (2015–2018). The data shows that Saturday and Sunday have the greatest number of UAS sightings, while counts on Monday, Tuesday, Wednesday, and Thursday are similar to one another.

Table 1. Sample Unmanned Aircraft System (UAS) Sighting Reports

Date	City	State	Summary	Info from FAA OPS	Alert	Number
1/23/15 12:38 p.m.	Ormond Beach	FL	OMN tower reported GA ACFT sighted a UAS 200' above the aircraft which was at 1000' In Omn's pattern. The UAS was on the right downwind leg of Ry17 at Omn. Den and Roc notified. Voulisa County Sheriff's Office Notified.	Ormond Beach, Fl/Uas incident/1238E/Daytona APRCH advised Cessna C172 at 1,000 feet on right downwind runway 17 reported UAS at 200 feet above heading S Bound. No conflicts reported. Volusia County PD.	MOR alert for DAB	DAB-M-2015/01/23-0002
1/23/15 6:30 p.m.	Minneapolis	MN	EPPD notified ATC of a UAS flying approx. 1.5 miles north of field with green and white lights. I did not see anything on radar or out the windows. EPPD Called Fcmt on the tower dispatch radio advising that they now know where it landed. Den notified and advised to send an officer to speak with the 'pilot' of the UAS. No traffic was affected.	Minneapolis, Mn/UAS Incident/1830C/FCM Atct advised Leo Reported UAS operating 1.5 N FCM. No impact to ops. UAS not observed from tower or on Radar. Eden Prairie PD investigating.	MOR alert for FCM	FCM-M-2015/01/23-0002
1/25/15 5:55 p.m.	Seattle	WA	ACFT advised tower they had observed a UAS just east of the Double Tree Inn in vicinity of 188th St And International Blvd at 100 ft. During the Seattle Seahawks sendoff parade. Approximately 10 minutes later the pilot observed the UAS at 800 ft. From the same location flying westbound in the direction where the team buses were heading (Delta Hanger). Pilot advised UAS was white in color and approximately 100–200 ft below his ACFT. OS notified local law enforcement via 911 on land line (King County Sheriff).	SEA/UAS Incident/0955P/Sea Atct received report from Robinson R22 at 900 feet of a white UAS flying 100–200 feet below, heading westbound from the downtown area toward the sea airport delta hanger. King County Sheriff notified.	MOR alert for SEA	SEA-M-2015/01/25-0001

Sundays had the greatest number of sightings with 1,621, and Saturdays had the second most sightings with 1,366.

Figure 5 illustrates the number of UAS sightings for the four states with the greatest number of sightings from 2015 to 2018 (California, Florida, New York, and Texas). A scatter plot for each year was created to illustrate the number of sightings for each month for these states.

Table 2 lists the ten locations in the U.S. with the highest number of mandatory occurrence report (MOR) alerts. All of these locations are either terminal radar approach control facilities or airports. The location with the highest number of MOR alerts is New York terminal radar approach control facility (TRACON) with 415 alerts, and the second-highest number is Southern California TRACON.

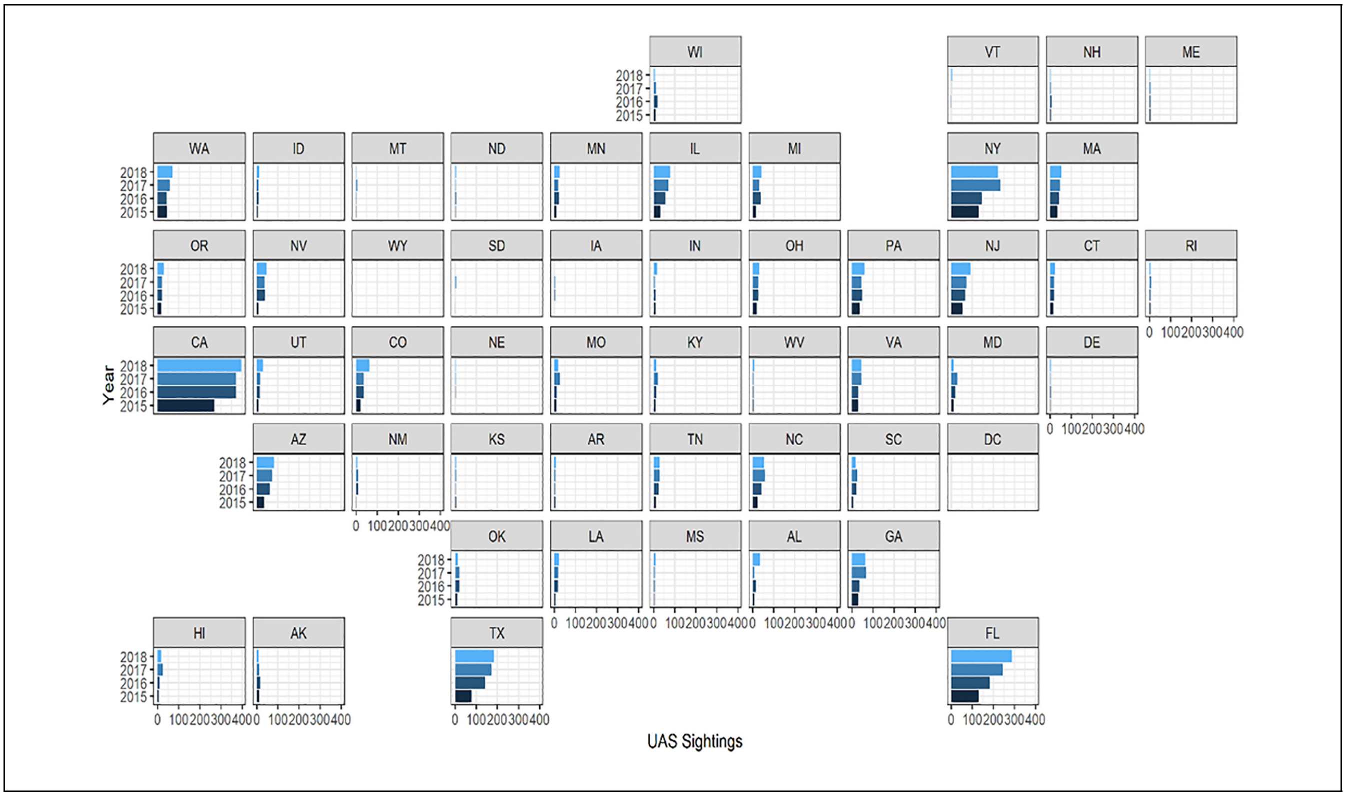

Figure 1. Unmanned aircraft system (UAS) sightings by state.

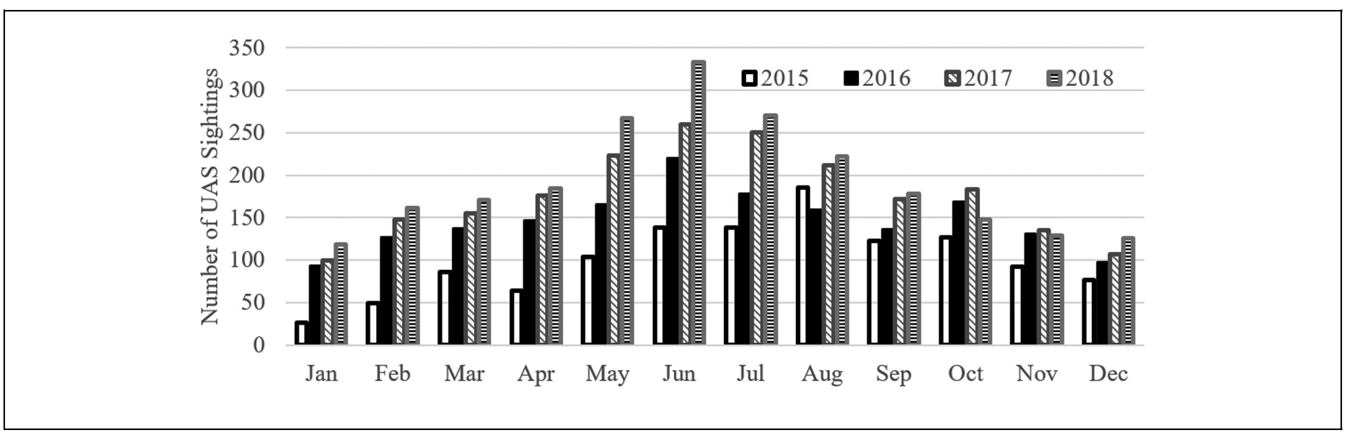

Figure 2. Number of unmanned aircraft system (UAS) sightings by month by year.

The exploratory data analysis (EDA) provided a brief of the nature and patterns of UAS-related incidents in the U.S. that occurred from 2015 to 2018.

Latent Dirichlet Allocation (LDA)

In 2003, Blei et al. introduced the LDA model to overcome the concerns associated with the probabilistic latent semantic analysis (PLSI) model (*15–17*). Improving on the PLSI model, it applies a K-dimensional latent random variable. The LDA model is one of the most popular probability topic models; however, it requires additional thorough assumptions of text generation than other models (*18*). The parameter space of this model is even simpler than PLSI model parameters. Additionally, the size of the parameter space is not relevant to the number of training documents in LDA; thus, it is a hierarchical model with a more stable structure, avoiding any overfitting situation (*18*, *19*).

Suppose a corpus (a large and structured set of texts) has three topics: UAS sightings information for three

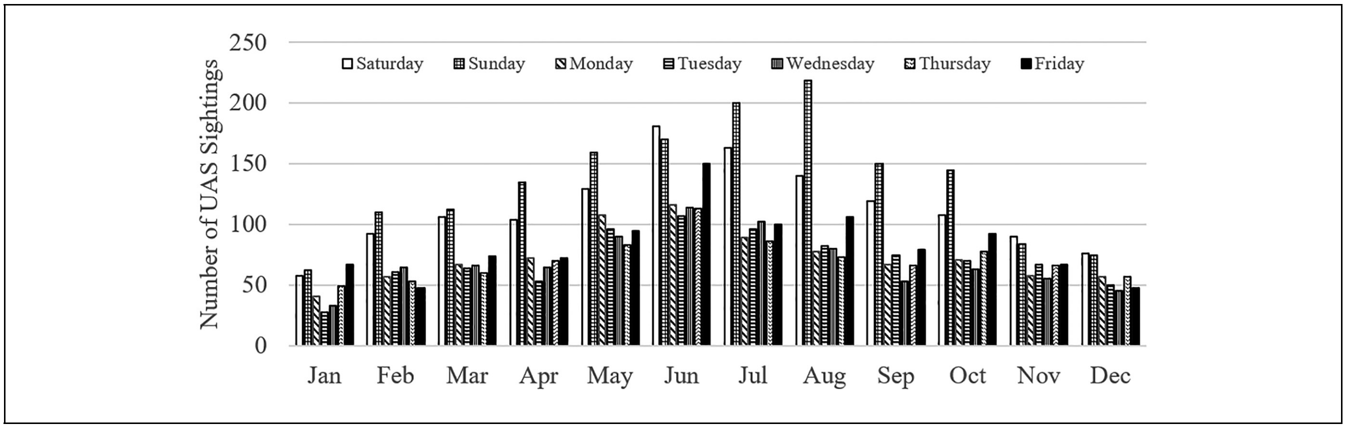

Figure 3. Number of unmanned aircraft systems (UASs) by month by day of the week.

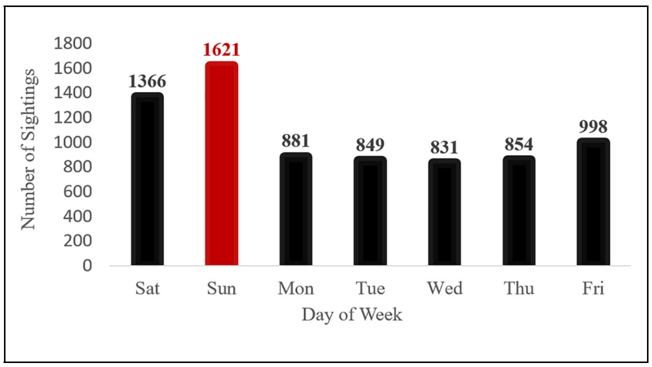

Figure 4. Number of unmanned aircraft system (UAS) sightings by the day of the week.

locations such as beach, city, and roadways. A document describing beach-associated UAS sightings may also contain information about both city- and road-related UAS sightings. Thus, in document generation, the topics will be chosen randomly first. However, the probability of choosing two topics encompassing city- and road-related UAS sightings will be higher for the beach-associated UAS sightings document. After selecting N words continually, a document is created. The LDA model is a three-tier Bayesian model and a probabilistic generative model that models discrete sets of data. This study applied the algorithms used in Girolami and Kabán to perform the analysis (*18*). Several recent transportation studies applied LDA in determining topics from unstructured textual contents (*20–22*). The current framework is based on the following assumptions:

1. *Document-topic association:* For each document d in corpus $\mathcal{D}$, a random variable $\theta_d \in \mathbb{R}^K$ is drawn from the Dirichlet distribution given by $\theta_d \sim Dir(\alpha)$. Here θ_d is the relative proportion of K topics in a given document.
2. *Word count:* For each topic k, a random variable $\beta_k \sim Dir(\eta)$ is drawn to postulate the distribution of terms in that specific topic.
3. *Topic-word association:* For each term t in document d, a topic $z_t \sim Mult(\theta_d)$ can be drawn from a multinomial distribution with θ_d prior and a scaled word frequency $tf_t \sim Mult(\beta_{z_t})$ from the multinomial distribution.

Results

Bigram Analysis

The research team also used the unstructured event reports to perform text mining for identifying patterns, trends, and hidden measures. All text mining steps (for example, token development, stop word removal, punctuation removal) have been conducted to perform the keyword extraction process. Lemmatization (consider similar words or word fragments as a consistent word) was also performed before the n-gram (contiguous sequence of n words from a document) analysis and topic modeling. A group of two words in the sequence is termed as a bigram. Figure 6 illustrates the top 15 most frequent bigrams from four documents representing the years 2015, 2016, 2017, and 2018; for example, all UAS sighting reports are merged into one document to provide the 2015 document. Before conducting topic modeling, the research team must define the clusters. The bigrams from the year-based clusters show similar bigrams, with a few exceptions. The top words such as "quad copter," "aviation unit," "action unknown," "action notification," and other airport-relevant keywords indicate that the UAS occurrences are associated with human interactions and improper decision-making. From the exploration of sighting reports, it is usually found that the local police authority was asked to perform an investigation. However, the investigation reports

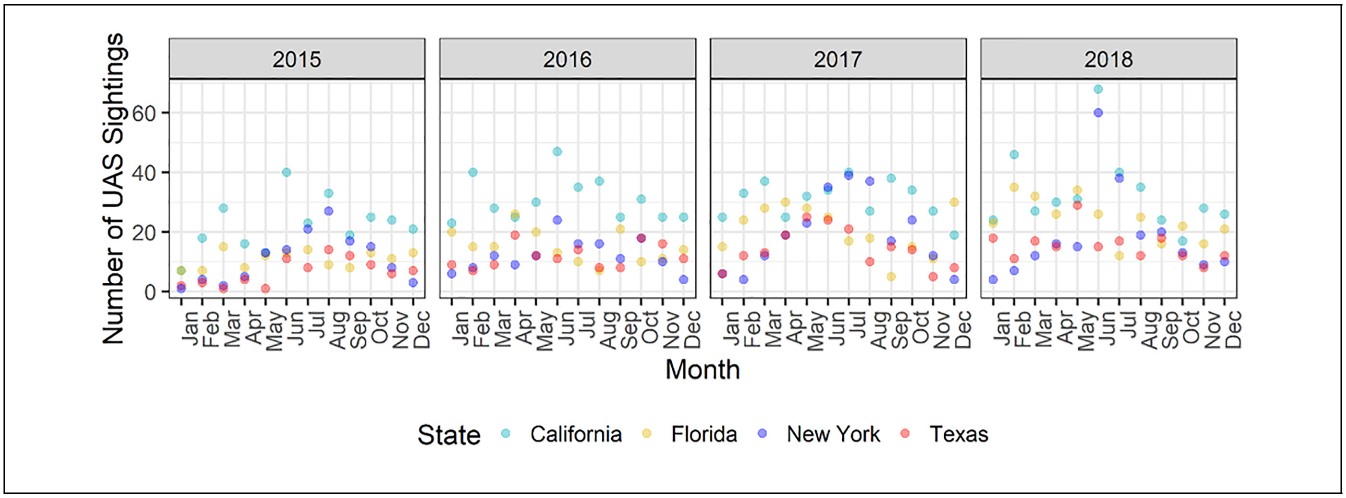

Figure 5. Number of unmanned aircraft system (UAS) sightings by month (CA, FL, NY, and TX).

Table 2. Top 10 Locations with High Mandatory Occurrence Report (MOR) Alerts

Location	MOR alerts
New York TRACON	415
Southern California TRACON	285
Northern California TRACON	159
La Guardia Airport, New York	145
Potomac Consolidated TRACON, Virginia	129
Chicago TRACON, Illinois	123
Los Angeles International Airport, California	118
Miami International Airport, Florida	107
Houston TRACON, Texas	99
John F. Kennedy International Airport, New York	98

Note: TRACON = terminal radar approach control facility.

are not publicly available. The bigram analysis reveals a need for additional UAS-sighting-related stop words to determine more intuitive results. Future studies can mitigate the current research gap. Based on the preliminary exploration, the topic modeling was conducted based on the document clusters of UAS sighting reports for each month, totaling 48 documents.

Topic Models

As the unstructured event documents provide detailed information of the associated event, it is important to develop topic models to see the trends and patterns. This study applied LDA to develop per-topic-per-word probabilities (β). For example, the term "heading" has a 0.01 probability of being generated from Topic q, but it has a 0.125 and 0.0095 probability of being generated from Topic 2 and Topic 4, respectively. Figure 7 provides users with a visual representation of ten topics that are extracted from the documents. The number of topics is limited to 10 for interpretation purposes only. A majority of the words in each topic indicate interactions between humans and UASs for different contexts. The most common words in Topic 1 are "heading," "police," "vicinity," "nypd," and "unidentified;" this suggests a topic related to unidentified UAS sighting incidents in New York. Topic 5 and Topic 9 also indicate location-specific UAS sightings. The second topic has keywords like "heading," "beach," and "vicinity," which may represent beach or open space UAS sightings.

Topic 7, Topic 8, and Topic 9 are also related to the beach or open space. Almost all of the topics contain "copter" or "quadcopter" in the list of keywords (exception: Topic 1). It indicates that UASs or "quadcopters" are identifiable during the manned air travels, as UASs fly either near or below their flight paths. The findings are in line with Gettinger and Michel (*13*). In many cases, local law enforcement authorities were involved in the investigation, and the operators of the UASs involved in the sighting incidents are unidentified. There is a need for additional measures to trace the unidentified UAS operators. As 10 topic models cannot provide a comprehensive picture of the topical distribution and occurrence of words in between topics, there is a need for a flexible parameter tool to generate additional insights. The following section provides additional flexibility in identifying topical trends in these 7,400 summary reports.

LDA Visualizations

Applying LDA to 7,400 documents will generate many topic combinations, which are modeled as distributions across thousands of terms. Limiting the number of topics

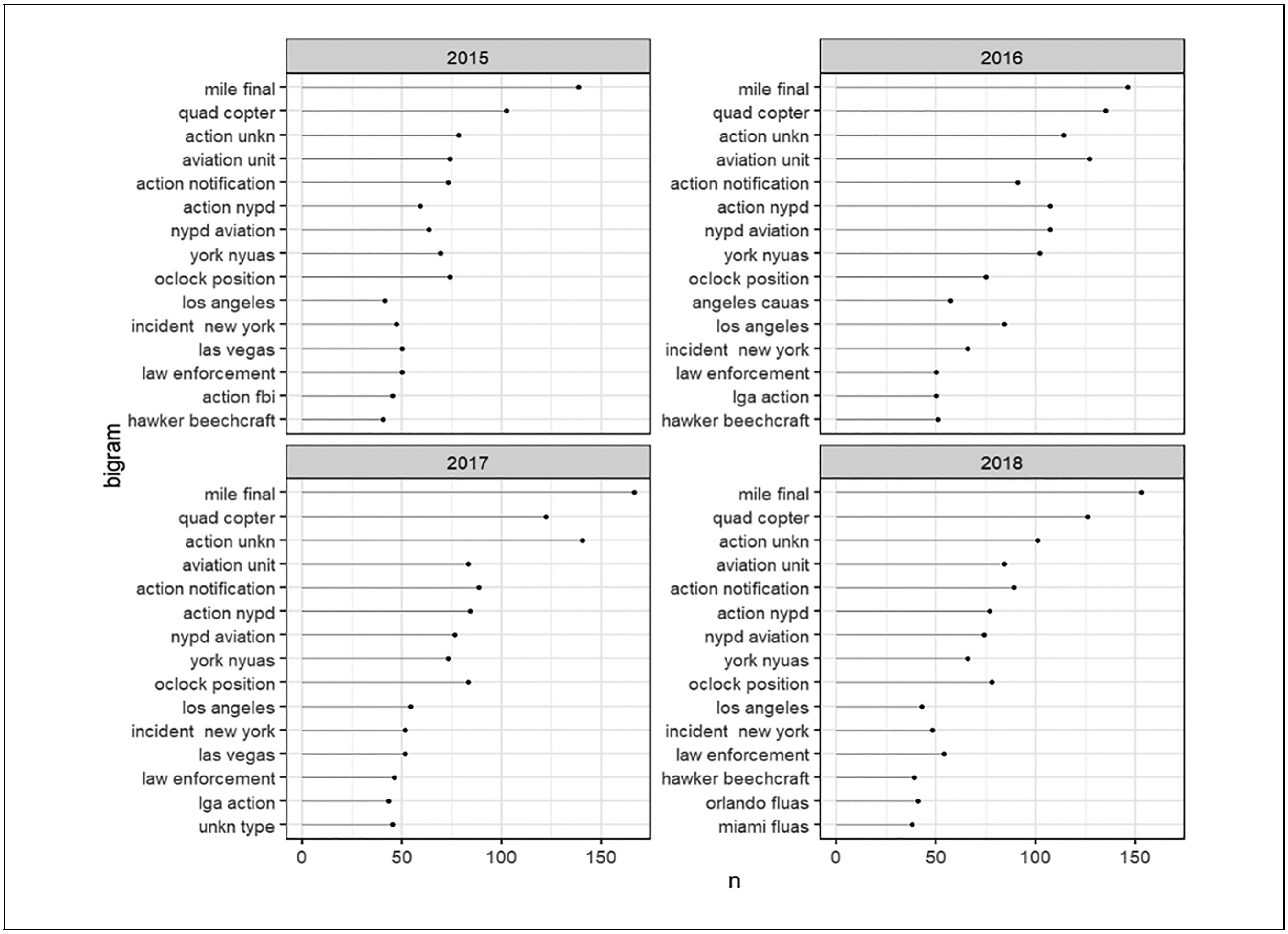

Figure 6. Top 15 bigrams by year.

to a limited count does not serve the research purpose. Research has shown that interactivity can alleviate some of the challenges. LDA visualization can mitigate the current research gap. The opensource R package LDAvis package was used to develop interactive LDA models (*20*, *23*). As a part of this study, an interactive webtool has been developed using the narrative texts used in 7,400 reports. It contains options for selecting the relative metric and number of topics (*24*).

Figure 8 shows the screenshot of this interactive visualization that is comprised of two sections:

- The left side of the visualization provides a global perspective of the developed topic models in relation to principal component analysis (PCA). PCA is a dimension reduction method to show the relevance of topics based on the relative distances in a two- or three-dimensional PCA plot. The topics are plotted as circles with centroids in the PCA plot with dimension 1 and dimension 2. The centers of these circles are based on the computed distance between topics by projecting the inter-topic distances onto two dimensions using multidimensional scaling (*25*). Each topic's overall prevalence is then encoded using the circle radius (e.g., Topic 1 shows the highest marginal topic distribution). Topic 1, Topic 2, Topic 4, and Topic 6 are on the lower right side of the PCA plot. Three topics (Topic 2, Topic 4, and Topic 6) overlap with each other. It indicates that these topics have some common keywords. Hovering over the circle shows red bar plots to the right. Additional insights can be drawn from the bar plot keyword distribution of each topic. The presence of Topic 5 is on the upper right side. On the upper left side, there are two topics: Topic 3 and Topic 10. The other three topics (Topic 7, Topic 8, and Topic 9) are on the lower left side.
- The right section displays a horizontal bar chart with 30 keywords per topic. The bars represent the representative 30 keywords by topic on the left, based on which topic is selected or hovered. The overlapped bars (blue and red, respectively) show

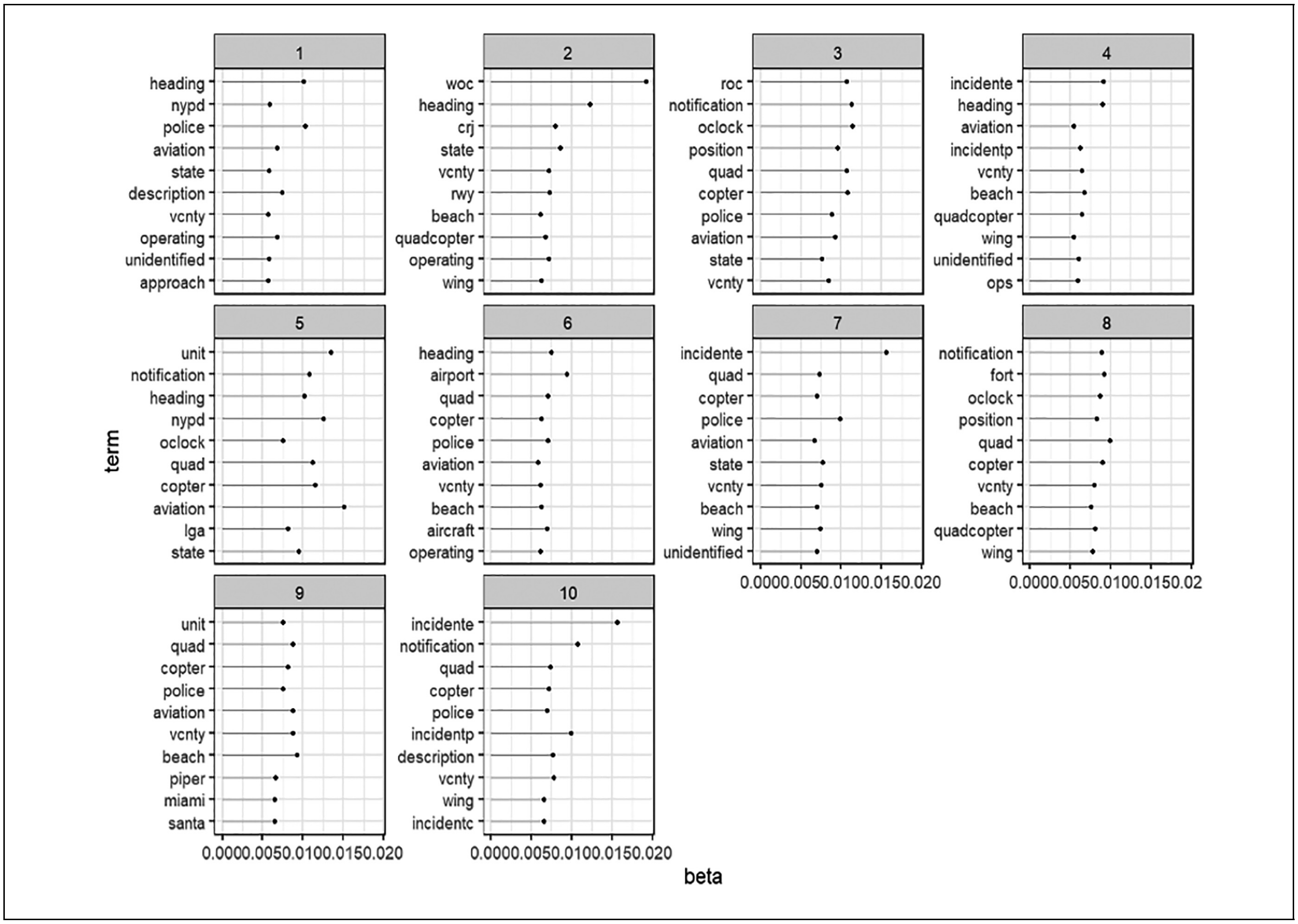

Figure 7. Top 10 topic models with 10 keywords with high beta.

both the corpus-wide and topic-specific counts of a top keyword.

- On the top right side, the relevance metric (λ) can be selected by dragging the bar from 0 to 1. After hovering a topic on the left, the red bar plots will be selected. Users can then drag the bar in between 0 to 1 to see the interactive changes of the bar plots according to their respective frequency changes.

This tool allows users to explore multiple combinations of topic-term relationships with flexible options. However, it is often very difficult to explain interactive tools in a two-dimensional plot. Interested users can explore the interactive tool (http://subasish.github.io/pages/uas_lda01) to explore the associated topics and their interrelations in a PCA plot.

Conclusion

UAS technology has previously been used almost exclusively by hobbyists and military personnel. Since 2012, however, UAS technology has transitioned to a broader range of uses, such as commercial and scientific applications; military and recreational use has also increased in recent years. As the application of UAS has expanded, it has placed a strain on existing NAS regulations. This growth has presented challenges in ensuring that future UAS use is integrated smoothly and coordinates with the current manned aircraft and air traffic management systems to promote safety and efficiency. As new technologies continue to emerge, the current study can help researchers gain an understanding of these interactions.

The descriptive statistics of this study provide a comprehensive view of the patterns and factors in UAS sighting events. The performance metric developed in this study quantifying the keywords in different topics (via LDA visualization tool) encompasses potential risk factors that UAS might cause. The findings also show that some UAS violations are untraceable because of the current gaps in tracing the offenses.

The frequencies of the keywords in the form of events can help researchers understand the areas of risk and

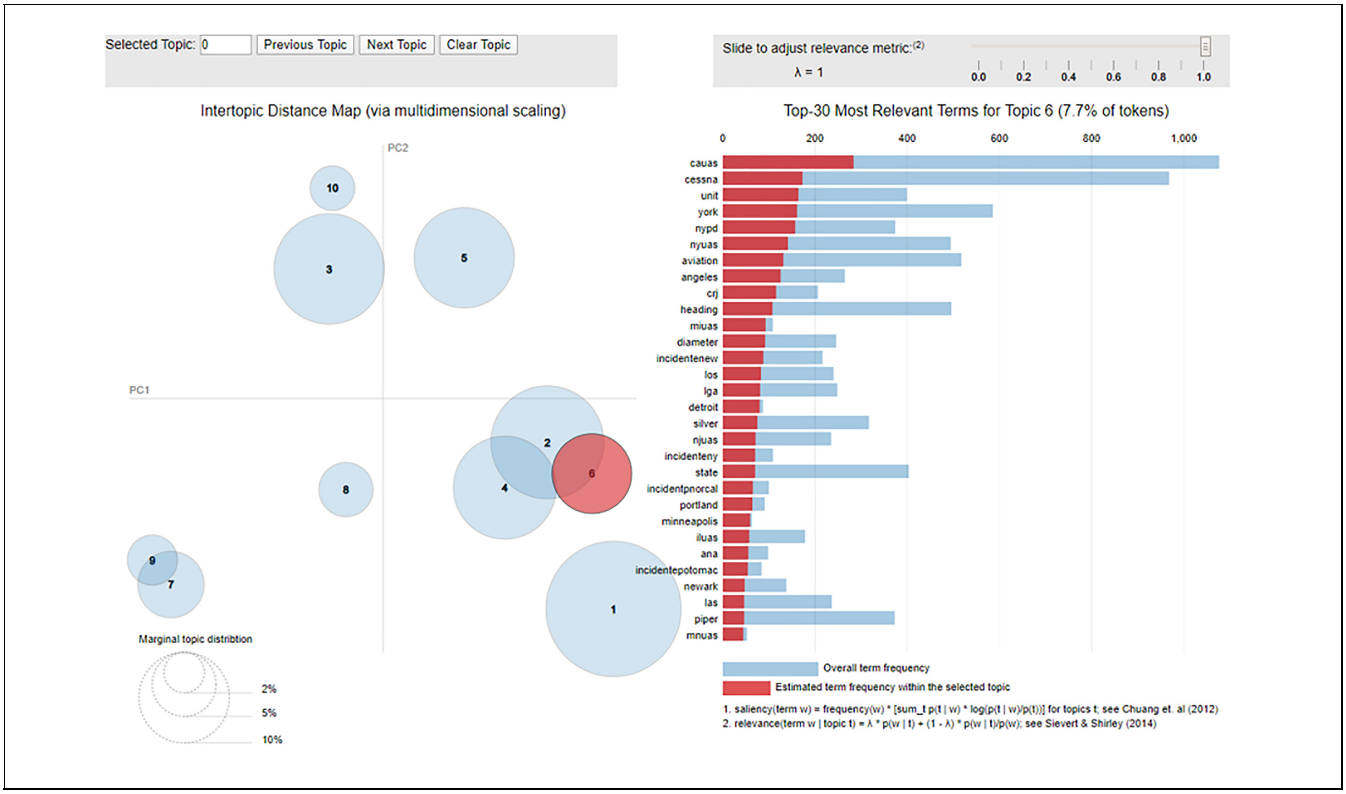

Figure 8. Interactive visualization of Latent Dirichlet Allocation (LDA) topic models.

develop practices to avoid UAS-related incidents and make the airspace safer for everyone. Insights from this study can help authorities design policies to trace unauthorized UAS sightings and reduce potential hazards. The overall safety improvement will depend on multiple stakeholders and their collaborative efforts. This study has three contributions: (1) it provides a comprehensive overview of UAS sightings in the U.S. from the past 4 years; (2) it ascertains the meaning of report narrative contents by presenting an extensive list of keywords, bigrams, and topics; and (3) the study provides an interactive web platform in which users can explore various topic models with different relevance matrices. In the future, this study will provide valuable insights for future UAS researchers. It will also provide the FAA, state DOTs, industry partners, and other UAS stakeholders with valuable insights as they continue to integrate UAS into the NAS.

Acknowledgments

The author would like to thank three anonymous reviewers for their comments. The paper has been improved greatly because of their valuable suggestions.

Declaration of Conflicting Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author received no financial support for the research, authorship, and/or publication of this article.

References

1. Kuhn, K. *Small Unmanned Aerial System Certification and Traffic Management Systems*. No. PE-269-RC. RAND Corporation, Santa Monica, CA, 2017, p. 19.
2. Itkin, M., M. Kim, and Y. Park. Development of Cloud-Based UAV Monitoring and Management System. *Sensors*, Vol. 16, No. 11, 2016, p. 1913.
3. Reddy, L. B., and D. DeLaurentis. Opinion Survey to Reduce Uncertainty in Public and Stakeholder Perception of Unmanned Aircraft. *Transportation Research Record: Journal of the Transportation Research Board*, 2016. 2600: 80–93.
4. Pankok, C., and E. J. Bass. A Decadal Revisiting of the Assessment of Pilot Control Interfaces for Unmanned

Aircraft Systems. Presented at the Human Factors and Ergonomics Society Annual Meeting, Los Angeles, CA, Vol. 61, No. 1, 2017, pp. 63–67.
5. Williams, K. W. *Documentation of Sensory Information in the Operation of Unmanned Aircraft Systems*. Federal Aviation Administration, and Federal Aviation Administration, Washington, D.C., 2008, p. 57.
6. Williams, K. W. *An Assessment of Pilot Control Interfaces for Unmanned Aircraft*. Federal Aviation Administration, and Federal Aviation Administration, Washington, D.C., 2007, p. 18.
7. Tobór, D., J. Barcik, and P. Czech. Legal Aspects of Air Transport Safety and the Use of Drones. *Scientific Journal of Silesian University of Technology: Series Transport*, Vol. 97, 2017, pp. 167–179.
8. Khan, M. A., W. Ectors, T. Bellemans, D. Janssens, and G. Wets. UAV-Based Traffic Analysis: A Universal Guiding Framework Based on Literature Survey. *Transportation Research Procedia*, No. 22, 2017, pp. 541–550.
9. Yuan, X., Z. Feng, W. Xu, W. Ni, J. A. Zhang, Z. Wei, and R. P. Liu. Capacity Analysis of UAV Communications: Cases of Random Trajectories. *IEEE Transactions on Vehicular Technology*, Vol. 67, No. 8, 2018, pp. 7564–7576.
10. Lennertz, T., A. L. Sparko, K. Cardosi, A. Yost, A. Kendra, J. Lu, and T. Sheridan. Pilots' Estimation of Altitude of a Small Unmanned Aircraft System (SUAS). *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 62, No. 1, 2018, pp. 49–53.
11. Goodchild, A., and J. Toy. Delivery by Drone: An Evaluation of Unmanned Aerial Vehicle Technology in Reducing CO^2 Emissions in the Delivery Service Industry. *Transportation Research Part D: Transport and Environment*, Vol. 61, No. Part A, 2018, pp. 58–67.
12. Battista, A., and D. Ni. A Comparison of Traffic Organization Methods for Small Unmanned Aircraft Systems. *Transportation Research Record: Journal of the Transportation Research Board*, 2018. 2672: 18.
13. Gettinger, D., and A. Michel. *Drone Sightings and Close Encounters: An Analysis*. Center for the Study of the Drone, Bard College, 2015.
14. U.S. GAO. *Small Unmanned Aircraft Systems: FAA Should Improve Its Management of Safety Risks*. U.S. GAO, Washington, D.C., May 2018.
15. Blei, D. M. *Probabilistic Models of Text and Images*. University of California, California, 2004.
16. Blei, D. M., A. Y. Ng, and M. I. Jordan. Latent Dirichlet Allocation. *The Journal of Machine Learning Research*, Vol. 3, 2003, pp. 993–1022.
17. Blei, D. M., and M. I. Jordan. Variational Inference for Dirichlet Process Mixtures. *Bayesian Analysis*, Vol. 1, No. 1, 2006, pp. 121–144.
18. Girolami, M., and A. Kabán. On an Equivalence Between PLSI and LDA. *SIGIR '03 Proc., 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Toronto, Canada, 2003, pp. 433–434.
19. Masada, T., S. Kiyasu, and S. Miyahara. Comparing LDA with PLSI as a Dimensionality Reduction Method in Document Clustering. *LKR'08 Proc., 3rd International Conference on Large-Scale Knowledge Resources: Construction and Application*, Tokyo, Japan, 2008, pp. 13–26.
20. Das, S., A. Dutta, and M. Brewer. Case Study of Trend Mining in Transportation Research Record Articles. *Transportation Research Record: Journal of Transportation Research Board*, 2020. 2674: 1–14.
21. Das, S., and A. Dutta. Twelve Years of Transportation Annual Meeting Hashtag: Implications for Networking and Research Trends. *Transportation Research Record: Journal of the Transportation Research Board*, 2021.
22. Das, S., and A. Dutta. Characterizing Public Emotions and Sentiments in COVID-19 Environment: A Case Study of India. *Journal of Human Behavior in the Social Environment*, 2020, pp. 1–14. https://doi.org/10.1080/10911359.2020.1781015
23. Sievert, C., and K. Shirley. LDAvis: A Method for Visualizing and Interpreting Topics. *Proc., Workshop on Interactive Language Learning, Visualization, and Interfaces*, Baltimore, Maryland, USA, June 27, 2014, pp. 63–70.
24. *Topic Model on UAS Sighting Narrative Reports*. http://subasish.github.io/pages/uas_lda01. Accessed December 2020.
25. Chuang, J., D. Ramage, C. Manning, and J. Heer. *Interpretation and Trust: Designing Model-Driven Visualizations for Text Analysis*. CHI, Austin, TX, 2012.