AMCIS 2022 Proceedings

Conference Theme Track - Innovative Research Informing Practice

Aug 10th, 12:00 AM

Coastal Resilience Decision Making with Machine Learning

Carol Lee
Northeastern University, carol.lee002@umb.edu

YoungHo Yoon
University of Massachusetts Boston, youngho.yoon001@umb.edu

Pratyush Bharati
University of Massachusetts, pratyush.bharati@umb.edu

Coastal Resilience Decision Making with Machine Learning

Emergent Research Forum (ERF)

Carol Lee
Northeastern University
Car.Lee@northeastern.edu

YoungHo Yoon
University of Massachusetts Boston
YoungHo.Yoon001@umb.edu

Pratyush Bharati
University of Massachusetts Boston
Pratyush.Bharati@umb.edu

Abstract

Our research aims to understand how social data can be integrated with climate data using machine learning for coastal resilience decisions. Although data analytics techniques have been adapted for decision models, incorporating unstructured data is a challenge. We adapt a design science research approach to develop a coastal resilience decision model that can accommodate various sets of climate criteria and social attributes to help us understand coastal risks in communities vulnerable to coastal hazards. We collected social data from environmental groups and individuals and conducted an exploratory social media data analysis on coastal resilience in the greater Boston, U.S., area. We employ non-negative matrix factorization (NMF), a topic modeling technique, to extract human-interpretable topics from a preliminary dataset of 131 documents from 50 different accounts. The outcomes of this research can help community members and policy makers understand and develop robust sustainability and climate focused decisions.

Keywords

Machine learning, social media, coastal resilience, climate decision models, topic modeling.

Introduction

Globally, a billion people are projected to be at risk from coastal-specific climate hazards in low-lying cities (IPCC 2022). Our research aims to understand how social data can be integrated with climate data using data analytics and machine learning (ML) for coastal resilience decisions. More specifically, we aim to identify the key stakeholders, needs and actions of communities with ML approaches. Black, Indigenous and People of Color (BIPOC) communities have historically suffered disproportionate impact because of climate change and natural disasters.

We adapt a design science research (DSR) approach (Hevner et al. 2004; Peffers et al. 2007) to develop a coastal resilience decision model that analyzes climate resilience data sets, including social data from communities vulnerable to coastal hazards in the greater Boston, U.S. area, including many vulnerable BIPOC communities. We incorporate ML into a flexible decision model to assess coastal community needs and demands from social data. Decision makers, who take social and regulatory factors into consideration, can make enhanced climate vulnerability assessments with ML-enhanced models. For instance, social media and sensor data were analyzed with supervised learning and natural language processing to help forecast storms and disaster risks (Harvey et al. 2019; Kankanamge et al. 2020).

Although data analytics techniques have been adapted for decision models, incorporating unstructured data, such as images and videos, to address theory and practice is a challenge (Bharati 2017, pp. 273-276). Social factors play an important role in decision making on sustainability (Carberry et al. 2019) and the role of IS is integral to solving climate change challenges (Seidel et al. 2017). AI, ML, and IS literature can facilitate environmental governance and improve organizational processes and individual practices (Enenkel et al. 2020; Nishant et al. 2020). For example, classification methods were applied to historical flood behaviors to understand flood resilience (Saraviet al. 2019). Community and climate factors' impact on agricultural vulnerabilities were measured using ML to improve resource management in Bangladesh's coastal regions (Jakariya et al. 2020). We plan to address these challenges by extending IS literature in ML and climate change decision models to further our understanding of climate change impact. Our decision model, enhanced with social data, can accommodate various sets of climate criteria and social attributes to help us understand coastal risks on vulnerable communities. The outcomes of this research can help community members and policy makers understand and develop robust sustainability and climate focused decisions.

Research Design

This research adapts a DSR approach for designing and evaluating an information technology (IT) artifact, broadly defined as "constructs (vocabulary and symbols), models (abstractions and representations), methods (algorithms and practices), and instantiations (implemented and prototype systems)" (Hevner et al. 2004). In alignment with DSR methodology (Peffers et al. 2007), our process is designed to be flexible with an outcome-oriented evaluation for coastal resilience decision making. The research design includes the development of a coastal resilience decision model that employs ML for analyzing social data from coastal communities alongside climate observations (see Figure 1).

We applied Non-Negative Matrix Factorization (NMF), one of the prominent topic modeling techniques introduced by Lee and Seung (1999), to extract human-interpretable topics from the text that are eventually clustered (Casalino et al. 2018). It is known as the prominent dimensionality reduction technique for text data which returns low-rank factor matrices that uncover the underlying structures and/or patterns in the data (Lee and Seung 1999). Specifically, NMF factorizes the original high-dimensional text data matrix into two low-rank matrices that represent the intensity of terms per topic (Xu et al. 2003). It successfully extracts meaningful keywords that are semantically associated with each other. NMF is known to produce improved outcomes for sparse and short text data like microblogging sites due to its ability to extract the compact low-rank feature matrix that uncovers the most important features for each topic (Yan et al. 2013). This technique is suitable for our study that intends to extract underlying key voices of communities from a large data corpus.

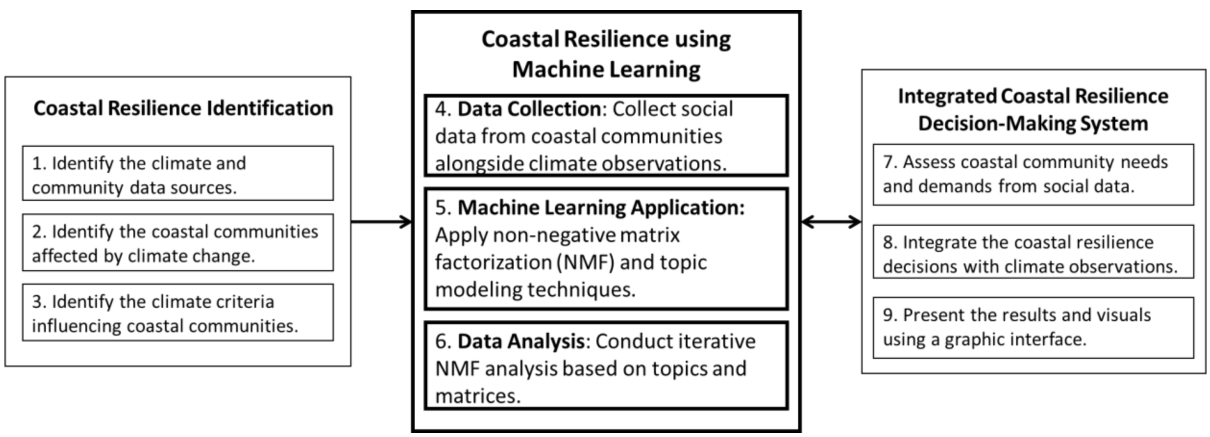

Figure 1. Coastal Resilience with Machine Learning Research Design

Methodology

Data Collection

To understand coastal resilience and climate justice discussions from community members and local organizations, we have collected social data from several social media accounts and pages of environmental groups including non-government organizations (NGOs), social movement organizations (SMOs), and from posts by individuals, all active in the greater Boston area. We used the groups' names and climate-related keywords to collect data from their public Facebook pages and the Twitter accounts. Our preliminary dataset includes 131 documents, which were mostly created in 2021, from 50 different accounts.

Data Analysis

We employ NMF, introduced by Lee and Seung (1999), to extract human-interpretable topics from the social data using the NMF module in the scikit-learn package (Cichocki and Phan 2009) and NLTK package. We assign an id and text to each document to identify the semantic relationship between words and extract the key topics for all the documents. The framework of this study is based on a three-stage process (Casalino et al. 2018). The first stage is devoted to dataset creation that transforms a collection of raw data into a preprocessed dataset in a matrix V (Visible Variables). Documents and terms are represented in columns and rows respectively, where values are assigned according to the weighting function of term frequency and inverse document frequency. Stop words, low information words prominent throughout the dataset, such as mass, Massachusetts, and Boston, are added to the process to avoid unnecessary term calculations.

In the second stage, our NMF model will automatically factorize the data matrix to extract human-interpretable topics by reducing the dimensionality of the dataset into two low-rank factor matrices, which are the document-topic matrix W (weights) and topic-term matrix H (hidden variables). In the W matrix, each row represents a document composed of unnormalized probabilities of topics and each column represents a semantic feature, consisting of a topic. Each value in the W matrix represents the aggregated weights of topic-related terms. In the H matrix, each row represents a topic, and each column represents a visible variable, that is, terms. In this exploratory analysis, we assigned the number of topics to five in consideration of the potential variability of the outputs from each run of our NMF model. In the last stage, we analyzed the results of the factorization to understand the clusters of terms. The result of the data matrix produced the top five topics consisting of the most frequent terms for each topic and their weights.

Preliminary Results

Five topics extracted from our analysis represent community voices by environment-related SMOs and NGOs in Boston. Our preliminary results suggest how terms in a group are matched with the action types and key issues in the social media data. We examined the document-topic matrix W to see the association between topics and accounts, and to check whether commonalities or disparities among the group are found. The matrix shows that certain community groups focus on different key climate-focused goals that are localized to the communities, but also have a common goal (see Table 1).

Topic #	Terms (with Weights)	Overall Topic	Action Type (Key Issue)
Topic 1	river (1.14), neponset (1.13), rain (0.44), month (0.37), avenue (0.34), dotnews (0.28), greenway (0.24), bank (0.24), massdcr (0.21)	Neponset river flooding	Prevention of coastal hazards
Topic 2	climate (0.75), change (0.33), justice (0.32), community (0.30), action (0.28), resilience (0.24), emission (0.16), law (0.16), baker (0.14), governor (0.14)	Climate justice law petitions	Long-term demand on carbon emissions
Topic 3	tide (0.99), flooding (0.30), cedar (0.25), grove (0.23), trail (0.21), noreaster (0.19), wind (0.18), park (0.18), surge (0.17), wave (0.17)	Frequently flooded regions warning	Short/long-term demand on coastal hazard warnings
Topic 4	flood (0.68), substation (0.54), noeastiesubstation (0.34), zone (0.27), community (0.23), injustice (0.22), fun (0.18), fuel (0.18), site (0.17), question (0.16)	North Eastie Substation ban protest/East Boston Coastal Resilience	Short-term demand for flooding and substation ban
Topic 5	water (0.60), leaf (0.50), storm (0.40), flooding (0.27), drain (0.25), pollution (0.24), sea (0.17), area (0.17), catch (0.16), basin (0.16)	Flooding prevention activities during storms	Prevention of floods

Table 1. Preliminary Results from NMF Analysis

The identified topics largely focus on alerting community members on coastal hazards, such as flood events, or encouraging them to join environment activities, such as short-term individual actions for flood prevention activities and to support petitions or protests (see table 2). The results show that three watershed associations, such as Charles River Watershed Association (@CharlesRiverWatershed) and Neponset River Watershed Association (@NepRWA), are key stakeholders in sharing coastal resilience-related posts on social media. On the other hand, some organizations and individuals tend to focus on a specific issue they are interested in, such as a ban on a community substation or coastal hazard warnings.

Data Source	Posted By	Social Media Post Excerpts	Topic #
Twitter	Organization (@NepRWA)	… we've had 4.7" of #rainfall … in Dorchester is at its highest water level since February 2013! …	Topic 1
Facebook	Organization (@Charles River Watershed)	… We are asking you to step up and ask the legislature to prioritize clean water and climate resilience today! Gov. Baker proposed using $800M of #ARPA funds … Instead, the M.A. House is proposing $225M …	Topic 2
Twitter	Organization (@NepRWA)	Warning - the #NeponsetGreenway trail is flooded at Granite Ave in Dorchester due to a combination of the high tide and storm surge. …	Topic 3
Twitter	Individual	… Don't add a substation in a flood zone next to jet fuel to the list of injustices. Cancel the hearings & reopen the question of need. #NoEastieSubstation	Topic 4
Twitter	Organization (@MysticMyRWA)	Help keep leaves out of storm drains and off the sidewalk … Leaves that get washed down the storm drain can cause pollution in our waterways and flooding in our neighborhoods.	Topic 5

Table 2. Coastal-related Social Media Examples from NMF Topics

Preliminary Conclusion

Our preliminary results show the association between topics and accounts, the commonalities or disparities between the organizations and community members. Certain community groups focus on different key climate-focused goals that are localized to the communities, but also have a common goal. First, organizations' own activity targets are comparably wide, and second, one of their practices is sharing the content of other activists and organizations. Furthermore, we found that individuals are more likely to create content on flood warnings and reports (topics 1 and 3) while topics are more dispersed in the environment organizations' content. Individuals have a propensity to share the short-term events and goals-related content, such as flood alerts and protests of environmentally harmful facilities. The climate-related content created by organizations range from short-term goals on coastal improvements by individuals and local organizations to long-term collective goals through legislative changes.

Limitations and Future Research

Due to the variability of NML algorithm approximation, the resulting weights slightly differ every time the function is executed. Although the number of topics is fixed to reduce the variability, the five topics limitation may not be sufficient to capture all the community voices in our exploratory dataset. Further data analysis will adapt selection rules for the number of topics. We also plan to analyze a larger volume of social media data to integrate with climate observations for developing a coastal decision model leveraging ML techniques. The further progress from this research can help community members and policy makers better understand robust sustainability and develop climate focused decisions.

REFERENCES

Bharati, P. 2017. "Big Data Is Not a Monolith," in *Book Reviews INFORMS Journal on Applied Analytics*, W. Shen (eds.) (47:3), pp. 273-276.

Carberry, E., Bharati, P., Levy, D. L., and Chaudhury, A., 2019. "Social Movements as Catalysts for Corporate Social Innovation: Environmental Activism and Adoption of Green Information Systems", *Business and Society* (58:5), pp. 1083-1127.

Casalino, G., Castiello, C., Del Buono, N., and Mencar, C. 2018. "A framework for intelligent Twitter data analysis with non-negative matrix factorization," *International Journal of Web Information Systems* (14:3), pp. 334-356.

Cichocki, A., and Phan, A. H. 2009. "Fast local algorithms for large scale nonnegative matrix and tensor factorizations," *IEICE transactions on fundamentals of electronics, communications and computer sciences* (92:3), pp. 708-721.

Collins, C., Dennehy, D., Conboy, K., and Mikalef, P. 2021. "Artificial Intelligence in Information Systems Research: A Systematic Literature Review and Research Agenda", *International Journal of Information Management* (60), 102383.

Enenkel, M., Brown, M. E., Vogt, J. V., McCarty, J. L., Bell, A. R., Guha-Sapir, D., Dorigo, W., Vasilaky, K., Svoboda, M., Bonifacio, R., Anderson, M., Funk, C., Osgood, D., Hain, C., and Vinck, P. 2020. "Why Predict Climate Hazards If We Need to Understand Impacts? Putting Humans Back into the Drought Equation", *Climatic Change* (162:3), pp. 1161-1176.

Harvey, J., Kumar, S., and Bao, S. 2019. "Machine Learning-Based Models for Assessing Impacts Before, During and After Hurricane Florence", in *2019 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 714-721.

Hevner, A. R., March, S. T., Park, J., and Ram, S. 2004. "Design science in information systems research," *MIS Quarterly* (28:1), pp. 75-105.

IPCC, 2022: *Climate Change 2022: Impacts, Adaptation, and Vulnerability*. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [H.-O. Pörtner, D.C. Roberts, M. Tignor, E.S. Poloczanska, K. Mintenbeck, A. Alegría, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem, B. Rama (eds.)]. Cambridge University Press. In Press.

Jakariya, M., Alam, M. S., Rahman, M. A., Ahmed, S., Elahi, M. M. L., Khan, A. M. S., Saad, S., Tamim, H. M., Ishtiak, T., Sayem, S. M., Ali, M. S., and Akter, D. 2020. "Assessing Climate-induced Agricultural Vulnerable Coastal Communities of Bangladesh using Machine Learning Techniques", *Science of the Total Environment* (742), 140255.

Kankanamge, N., Yigitcanlar, T., Goonetilleke, A., and Kamruzzaman, M. 2020. "Determining Disaster Severity Through Social Media Analysis: Testing the Methodology with South East Queensland Flood Tweets", *International Journal of Disaster Risk Reduction* (42), 101360.

Lee, D. D., and Seung, H. S. 1999. "Learning the parts of objects by non-negative matrix factorization," *Nature* (401:6755), pp. 788-791.

Nishant, R., Kennedy, M., and Corbett, J. 2020. "Artificial Intelligence for Sustainability: Challenges, Opportunities, and a Research Agenda", *International Journal of Information Management* (53), 102104.

Peffers, K., Tuunanen, T., Rothenberger, M. A., and Chatterjee, S. 2007. "A Design Science Research Methodology for Information Systems Research," *Journal of Management Information Systems* (24:3), pp. 45-77.

Saravi, S., Kalawsky, R., Joannou, D., Rivas Casado, M., Fu, G., and Meng, F. 2019. "Use of Artificial Intelligence to Improve Resilience and Preparedness Against Adverse Flood Events", *Water* (11:5), 973.

Seidel, S., Bharati, P., Fridgen, G., Watson, R. T., Albizri, A., Boudreau, M. C. M., Butler, T., Kruse, L.C., Guzman, I., Karsten, H., Lee, H., Melville, N., Rush, D., Toland, J., and Watts, S. 2017. "The Sustainability Imperative in Information Systems Research", *Communications of the Association for Information Systems* (40), pp.1, 3.

Xu, W., Liu, X., and Gong, Y. 2003. "Document clustering based on non-negative matrix factorization", in proceedings of *the 26th annual international ACM SIGIR conference on Research and development in information retrieval,* pp. 267-273.

Yan, X., Guo, J., Liu, S., Cheng, X., and Wang, Y. 2013. "Learning topics in short texts by non-negative matrix factorization on term correlation matrix", in proceedings of *the 2013 SIAM International Conference on Data Mining*, pp. 749-757.