AMCIS 2022 Proceedings

SIG DSA - Data Science and Analytics for Decision Support

Aug 10th, 12:00 AM

Text mining for classifying workplace severe injury events

David M. Goldberg
San Diego State University, dgoldberg@sdsu.edu

Nohel Zaman
Loyola Marymount University, nohel.zaman@lmu.edu

Text Mining for Classifying Workplace Severe Injury Events

Emergent Research Forum (ERF)

David M. Goldberg
San Diego State University
dgoldberg@sdsu.edu

Nohel Zaman
Loyola Marymount University
nohel.zaman@lmu.edu

Abstract

Occupational health and safety are significant issues for many organizations. Workplace accidents have significant consequences for employees, and they also cause significant disruptions to business processes. In this paper, we analyze accident narrative data collected by the Occupational Safety and Health Administration (OSHA) to classify events leading to injuries and body parts injured. These datasets are collected in textual form, which can create difficulties for nuanced statistical analyses. By creating high-performing text classifiers, we can automatically code these records and provide decision-makers with useful information. We use BERT word embeddings and several variants thereof to perform these classifications, finding accuracy of approximately 89 percent and 79 percent for event types and body parts respectively. These contributions will improve organizations' capacities to understand and act upon textual accident narratives.

Keywords

Occupational safety, text mining, word embeddings, machine learning

Introduction

Work-related injuries, diseases, and fatalities obtained significant research attention over the past decade. It is estimated that approximately 10 workers experience a work-related accident per second (Abdalla et al. 2017) and that 2.78 million deaths per year can be attributed to occupational accidents and workplace-related diseases (Päivi Hämäläinen et al. 2017). According to the U.S. Bureau of Labor Statistics report[1], there were 1,176,340 nonfatal injuries and illnesses that caused a private industry worker to miss at least one day of work in 2020, a figure which was 32.4 percent higher than in 2019.

Globalization impacts the structure of workplaces and occupational safety and health (OSH) practices. Additionally, poor occupational safety and health inflict substantial financial losses on employers and society. Estimates from the International Social Security Association (ISSA) suggest that costs associated with nonfatal workplace accidents alone equal approximately 4 percent of world gross domestic product (GDP) each year (Abdalla et al. 2017). Li et al. (2021) mentioned that incremental investment in scientific research and education related to occupation safety and health is beneficial in enhancing the work quality of workers and training safety professionals, and hence can effectively lessen workplace accidents pertaining to specific safety concerns. Mahalingam et al. (2020) conducted a systematic review of cost-benefit analyses related to unintentional injury prevention. The research finds that many preventative strategies are cost-effective; that is, they produce benefits whose financial value more than offsets the costs. However, targeting these safety investments in the areas of the most significant need is an ongoing challenge.

Previous research (Goh and Ubeynarayana 2017; Mahalingam et al. 2020) has proposed that investigating the distinct nature of past workplace accidents allows organizations to develop steps to reduce future accidents. However, these analyses are complicated by the textual nature of data collection. Many

[1] https://stats.bls.gov/news.release/osh.nr0.htm

organizations used unstructured textual narratives to describe workplace accidents, and thus it is difficult to draw statistical conclusions as to accident types or severities. As a result, multiple recent studies took the challenge in adopting machine learning techniques to textual data for the automated detection of accident reports prone to safety concerns (Goh and Ubeynarayana 2017; Goldberg 2021; Zhang et al. 2019). Organizations could rapidly investigate and improve their safety practices by utilizing the automated methods in classifying the safety concerns present in accident narratives. Additionally, government agencies could use this refined information to modify occupational safety standards.

In this paper, we consider textual narratives on severe workplace injuries compiled by the Occupational Safety and Health Administration (OSHA). Due to the length and detail of these narratives, there is potentially great value in text mining to classify and describe the safety concerns expressed in these textual narratives. For instance, in the recent literature, Goldberg (2021) collected a dataset of accident narratives and classified according to five dimensions: hospitalization, amputation, body parts, source, and event. The author used word embeddings-based text mining techniques to perform automated classifications of accident narratives into multiple dimensions. Previous studies such as Goh and Ubeynarayana (2017) have utilized similar datasets with accident narrative text classification methodologies. Although the past studies made extensive progress in automated classification of multiple dimensions, there is still opportunity for assessing additional dimensions of events. This study examines advanced text mining techniques on OSHA accident narratives, focusing on two main dimensions: event causing the injury and body parts injured. These techniques provide value to OSHA government agency and employers by rapidly prioritizing accident reports pertaining to 17 different event types and 18 different body parts. Compared to Goldberg (2021)'s research goal that focused on higher-level event and body part classifications, this study aims to automate the detection of specific events such as "compressed or pinched by shifting objects or equipment" or "contact with hot objects or substances", and many more, appearing in multiple accident events. In addition, we also aim to disambiguate the body parts injured on a fine-grained level, for example disambiguating leg injuries between hips, thighs, feet, etc. Our results demonstrate the capability of text mining in occupation safety area and offer tools for professionals to sort and highlight reports in need of the most crucial inspection for a diverse set of accident events and body parts injured.

Related Work

The occupational safety and health (OSH) literature is robust and considers many dimensions of workplace safety and accident prevention. Workplace accidents are known to be deleterious for both employees and organizations; in addition to financial losses, supplementary costs such as physical and psychological pain are significant in many situations (Feng et al. 2015). Ikpe et al. (2012) assessed the numerous costs related to workplace injury events, which can cascade beyond the immediate employee injured and also affect their social circles of friends and family.

Workplaces often employ preventative measures to avoid accidents. Feng (2013) explains a set of seven possible categories safety investments: safety staffing, safety equipment and facilities, compulsory trainings, in-house safety trainings, safety inspections and meetings, safety incentives and promotions, and safety innovation. Importantly, the choice of which measures to deploy varies significantly across organizations (Feng 2013). For example, if workplace accidents are driven by faulty or dangerous equipment, then the organization might prefer to invest in safety equipment and facilities and/or in safety inspections and meetings. On the other hand, if workplace accidents are driven more by employee error, then compulsory or in-house safety trainings may prove necessary. Thus, profiling the specific type of safety concerns is a critical component of effective safety investments.

Text data is unstructured in nature. Hence, to get any machine learning model trained on textual data, transforming each unstructured textual document into vectorized features is important. In a "bag-of-words" model, each document is converted such that each word pertains to a frequency (Zhang et al. 2010). Recent literature (Goh and Ubeynarayana 2017; Goldberg 2021) applied the information retrieval approach term frequency-inverse document frequency (Tf-idf), which reweights words based on how often they occur in a single document versus the corpus of all documents. Several recent studies have utilized word embeddings models, a recent innovation that uses deep learning to consider the interrelationships between words (Goldberg 2021; Zaman et al. 2021; Zhang et al. 2020).

A stream of literature used text mining to classify accident narratives from various industries such as construction, manufacturing, and mining. Recently, Goldberg (2021) applied text mining to a prelabeled dataset of ~50,000 OSHA accident narratives, classifying reports for whether they referred to various target classes. In contrast to Goldberg (2021)'s work, Goh and Ubeynarayana (2017) collected a manually labeled dataset of 1,000 OSHA accident narratives related to the construction industry and trained a series of machine learning models to perform classifications. The authors used a classification scheme by the Workplace Safety and Health Institute, defining the event type that affected the injury, such as "falls" or "exposure to chemical substances." A series of past studies used the same dataset, such as Zhang et al. (2019) and Zhang et al. (2020). A commonality in these studies is that each study attempts to create an automatic coding tool such that organizations can quickly analyze their data and understand the drivers of workplace accidents. This information could then be used to ensure effective targeting of investments in workplace safety.

Dataset and Methodology

Dataset

This study utilizes a dataset of severe injury narratives collected from the OSHA website (https://www.osha.gov/severeinjury). This dataset contains accident narratives that refer to the body part(s) injured and event types. In total, the dataset includes 66,699 reports covering the period between January 1, 2015, and July 31, 2021. This dataset represented the complete set of records available at the time of download. Per OSHA's taxonomy, our dataset contained 348 distinct event types and 124 distinct body parts. Some of these event types and body parts had extremely sparse representation, so we only considered event types and body parts for which 1,000 or more records were available. As a result, we consider 17 event types and 18 body parts. Thus, we filtered the dataset to 32,167 reports to focus on the event types and body parts with sufficient representation.

Methodology

In this paper, we use word embeddings to classify our text. These deep learning-based methods are trained on large volumes of text and use neural networks to model the interrelationships between words. Each model generates high-dimensional vector representations of text, and these vectors can then serve as features in classification models. Word embeddings are a recent innovation relative to traditional bag-of-words models, which consider whether certain words are present in text and at what frequencies, but without consideration of their order or semantic relationships. Word embeddings models account for the interrelationships between words, allowing for substantial performance.

In our paper, we consider Bidirectional Encoder Representations from Transformers (BERT), which is considered a state-of-the-art word embedding model (Devlin et al. 2018). BERT is pre-trained on billions of words from BooksCorpus and Wikipedia and is a general-purpose word embeddings model that performs well in many applications.

In addition to BERT, we also consider two BERT variants that are trained on more specialized datasets. Bidirectional Encoder Representations from Transformers for Biomedical Text Mining (BioBERT) (Lee et al. 2020) is a variant of BERT that is additional trained on PubMed data to improve its performance at classifying biomedical textual passages. We also consider SciBERT (Beltagy et al. 2019), which is trained with data from Semantic Scholar also focusing on the biomedical domain. These two variants may offer competitive performance given their training on more specialized datasets.

We trained each word embeddings model using a ten-fold cross-validated design. To evaluate the performance of the various models, we compute four commonly utilized machine learning metrics: accuracy, prevision, recall, and F1-score. Accuracy refers to the proportion of the records in the holdout set correctly classified. Precision refers to the proportion of records classified as the target class that actually refer to the target class. Recall refers to the proportion of records that actually refer to the target class that are correctly classified as the target class. Finally, F1-score is a harmonic mean of precision and recall.

Results

First, we consider the results of our analysis of the event type category. In Table 1, we present these results for each of our three deep learning word embeddings models: BERT, BioBERT, and SciBERT. We found that each of the models performed relatively well, all scoring between 0.76 and 0.79 on each metric. In our dataset, we consider 18 distinct body parts, and thus the classification task is quite difficult due to its multinomial nature, and these scores are quite accurate in that context. We found that the BERT model was the most accurate of the three, scoring at 0.79 on each metric, but the BioBERT and SciBERT models only performed slightly worse than the BERT model.

Model	Accuracy	Precision	Recall	F1-score
BERT	0.79	0.79	0.79	0.79
BioBERT	0.77	0.77	0.77	0.77
SciBERT	0.76	0.76	0.76	0.76

Table 1. Model performance on event type

Second, we consider the results of our analysis on the body parts category. We present the results of these analyses in Table 2. Our findings suggest that all three models, BERT, BioBERT, and SciBERT, performed with very similar accuracy (0.89 on each metric, identical to three decimal places). Considering the multinomial classification amongst 17 event types, these figures represent remarkable levels of performance. In addition, it is notable that there was not a substantial difference between accuracy, precision, recall, and F1-score. This suggests that the classification was not particularly stronger or weaker for certain event types, but rather it was relatively uniformly high performing.

Model	Accuracy	Precision	Recall	F1-score
BERT	0.89	0.89	0.89	0.89
BioBERT	0.89	0.89	0.89	0.89
SciBERT	0.89	0.89	0.89	0.89

Table 2. Model performance on body part

When considering our results, one limitation of our research design is that our metrics do not reward the models for predicting a class that is similar to the correct one but not an exact match. For example, in our body part classification, we delineate between hips, thighs, feet, etc. If a model misclassifies a record that truly refers to hips as referring to thighs, this is treated as completely incorrect, even though it is much closer to correct than misclassifying hips as fingertips. In consultation with occupational safety professionals, future work could develop a scheme for awarding some partial credit for near-correct classifications, and future models could then be tuned and optimized for this new scheme.

Conclusions

Our results are greatly promising for the use of word embeddings models to classify accident narratives in an occupational safety context. In this work, we utilized BERT, BioBERT, and SciBERT word embeddings models to classify accident narratives for the specific event type and body part to which they prefer. We found substantial performance for all three models, but overall, the BERT model seemed to offer the highest levels of performance.

We note that our study design has several limitations. First, the size of the available dataset limited the analyses we could perform. Although we downloaded all data available, some sparsely represented classes could not be reasonably modeled at this sample size. Future research may collection additional data to address this potential limitation. In addition, another potential limitation is that we tested several BERT variants, but we did not exhaustively examine every word embedding model nor every model architecture. We chose our models due to their strong performance in past work, but future research could expand upon our analyses to perform more exhaustive trials.

We are hopeful that these findings will help to spur additional research on using text mining to analyze accident narratives, providing vital occupational safety data for organizations. Government regulators and businesses alike can benefit from these results and use them in practice to modernize analyses of occupational safety.

REFERENCES

Abdalla, S., Apramian, S. S., Cantley, L. F., Cullen, M. R., Mock, C., Nugent, R., Kobusingye, O., and Smith, K. 2017. "Occupation and Risk for Injuries," *Injury Prevention and Environmental Health. 3rd. The International Bank for Reconstruction and Development/The World Bank*.

Beltagy, I., Lo, K., and Cohan, A. 2019. "SciBERT: A Pretrained Language Model for Scientific Text," *ArXiv preprint*.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. 2018. "BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding," *ArXiv preprint*.

Feng, Y. 2013. "Effect of Safety Investments on Safety Performance of Building Projects," *Safety Science* (59), pp. 28-45.

Feng, Y., Zhang, S., and Wu, P. 2015. "Factors Influencing Workplace Accident Costs of Building Projects," *Safety Science* (72), pp. 97-104.

Goh, Y. M., and Ubeynarayana, C. 2017. "Construction Accident Narrative Classification: An Evaluation of Text Mining Techniques," *Accident Analysis & Prevention* (108), pp. 122-130.

Goldberg, D. M. 2021. "Characterizing Accident Narratives with Word Embeddings: Improving Accuracy, Richness, and Generalizability," *Journal of Safety Research*.

Ikpe, E., Hammon, F., and Oloke, D. 2012. "Cost-Benefit Analysis for Accident Prevention in Construction Projects," *Journal of Construction Engineering and Management* (138:8), pp. 991-998.

Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C. H., and Kang, J. 2020. "BioBERT: A Pre-Trained Biomedical Language Representation Model for Biomedical Text Mining," *ArXiv preprint* (36:4), pp. 1234-1240.

Li, C., Wang, X., Wei, C., Hao, M., Qiao, Z., and He, Y. 2021. "Analysis of the Correlation between Occupational Accidents and Economic Factors in China," *International Journal of Environmental Research and Public Health* (18:20), p. 10781.

Mahalingam, M., Peterson, C., and Bergen, G. 2020. "Systematic Review of Unintentional Injury Prevention Economic Evaluations 2010–2019 and Comparison to 1998–2009," *Accident Analysis & Prevention* (146), p. 105688.

Päivi Hämäläinen, Jukka Takala, and Kiat, T. B. 2017. "Global Estimates of Occupational Accidents and Work-Related Illnesses 2017," Workplace Safety and Health Institute.

Zaman, N., Goldberg, D. M., Gruss, R. J., Abrahams, A. S., Srisawas, S., Ractham, P., and Şeref, M. M. 2021. "Cross-Category Defect Discovery from Online Reviews: Supplementing Sentiment with Category-Specific Semantics," *Information Systems Frontiers*, pp. 1-21.

Zhang, F., Fleyeh, H., Wang, X., and Lu, M. 2019. "Construction Site Accident Analysis Using Text Mining and Natural Language Processing Techniques," *Automation in Construction* (99), pp. 238-248.

Zhang, J., Zi, L., Hou, Y., Deng, D., Jiang, W., and Wang, M. 2020. "A C-BiLSTM Approach to Classify Construction Accident Reports," *Applied Sciences* (10:17), p. 5754.

Zhang, Y., Jin, R., and Zhou, Z.-H. 2010. "Understanding Bag-of-Words Model: A Statistical Framework," *International Journal of Machine Learning and Cybernetics* (1:1-4), pp. 43-52.