Institution of MECHANICAL ENGINEERS

Journal of RISK AND RELIABILITY

Editoral on special issue “Text mining applied to risk analysis, maintenance and safety”

Proc IMechE Part O:
J Risk and Reliability
2024, Vol. 238(5) 903–904
© IMechE 2024
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1748006X241280066
journals.sagepub.com/home/pio
Sage

Acquiring knowledge for Risk Analysis often requires the consultation of dense engineering documents, which can be exhaustive in practice. Such documents contain relevant information in text format that can be used by experts to characterize the system and postulate different hypotheses on hazards, accident scenarios, and system responses. In this context, text mining is an expanding research area, which has become popular in both academia and industry as an effective way to automatically extract information from large documents. The application of text mining for Reliability, Availability, Maintainability and Safety (RAMS) purposes may include the identification of patterns in maintenance services and accident investigations, or the development of automated risk analysis reports. The overall goal is to uncover useful knowledge from text reports through the application of linguistic, statistical, and machine learning techniques. Currently, deep learning techniques are intensely studied for developing solutions to different natural language processing problems and their use in support to risk analysis, maintenance, and safety seems promising.

This Special Issue contributes to the advancement of the field of “Text Mining applied to Risk Analysis, Maintenance and Safety.” We are delighted to present five high-quality papers that were published in the Special Issue. These papers explored a range of topics, including the use of data mining to uncover causation correlations, constructing, and analyzing knowledge graphs from maintenance work orders, automatic semantic knowledge extraction from electronic forms, using BERT to classify injury leave based on accident descriptions, and developing a framework for classifying the severity of road accidents. The problems addressed are notably complex, involving intricate data sets, requiring sophisticated methods, and considering the integration of diverse information sources.

Ma et al.[1] employed data mining technologies to analyze 285 maritime accident reports to uncover causation correlations among risk factors. Text mining is used to extract keywords to identify critical factors, and the FP-Growth algorithm reveals hidden association rules, which are then quantitatively evaluated. Findings indicate that combinations of factors may lead to accidents even in good navigation conditions. Consequently, stakeholders should systematically assess and mitigate these interconnected risk factors to prevent maritime accidents.

The study of Stewart et al.[2] addresses the challenge of extracting technical information from unstructured, jargon-rich maintenance work orders (MWO) and combining it with structured data fields. The solution provides an intuitive interface that visualizes historic asset data as a knowledge graph. It uses deep learning and annotated training data to automatically construct knowledge graphs from unstructured text and structured data. Tested on industry-provided MWO and delay accounting data, these tools help reliability engineers efficiently analyze historic asset data for failure modes, maintenance strategies, and process improvements. The source code is available on GitHub under the Apache 2.0 License.

Wu et al.[3] proposed a supervised computer vision model to decompose PDF forms into nested microtables, which are then processed using a customizable rule bank to extract meaningful table content and semantic relationships. Demonstrated on an industry dataset of 37 maintenance procedure documents, the model successfully handled 373 pages and 1016 unique microtables. A web application, EMU (Extracting Machine Understandable Semantics from Forms), showcases how data from these tables can be automatically extracted and stored in JSON format. This method enables machine-automated search and data extraction at scale, crucial for maintenance and other procedural documentation.

Ramos et al.[4] explored the use of Natural Language Processing (NLP) techniques to analyze accident investigation reports for predicting injury leave occurrences. By applying Bidirectional Encoder Representations from Transformers (BERT) to text data and combining it with numerical and binary variables from the reports, the study inputs this data into a Multilayer Perceptron (MLP) to predict accident leave events. The methodology, tested on reports from an actual hydroelectric power company, achieved a reasonably high median accuracy. The study also specifically discusses reports with high and low prediction accuracy, highlighting the value of accident reports in supporting safety-related decisions.

Finally, in Valcamonico et al.,[5] a framework is developed combining NLP and Machine Learning (ML) to automatically classify road accident reports to aid experts in road safety analysis. The work compares two textual representation models, Hierarchical Dirichlet Processes (HDPs) and Doc2vec, alongside three ML classifiers: Artificial Neural Networks (ANNs), Decision Trees (DTs), and Random Forests (RFs). Applied to road accident reports from the US National Highway Traffic Safety Administration, the combination of HDP topic modeling and RF classification is found to provide the best balance between classification accuracy and explainability of the results.

To conclude, we believe that the papers selected for this Special Issue exemplify the innovative approaches for smart extraction of knowledge from different bodies of text and their application to risk analysis, maintenance, and safety.

In closing, we thank all the authors for their contribution and the expert reviewers for their work.

Editors

Márcio das Chagas Moura
CEERMA – Center for Risk Analysis, Reliability and Environmental Modeling, Department of Management Science, Universidade Federal de Pernambuco, Recife, Brazil

Piero Baraldi
Politecnico di Milano, Dipartimento di Energia, Italy

Enrique López Droguett
Department of Civil and Environmental Engineering, Garrick Institute for the Risk Sciences, University of California – Los Angeles, CA, USA

Enrico Zio
Politecnico di Milano, Dipartimento di Energia, Italy

ORCID iDs

Márcio das Chagas Moura https://orcid.org/0000-0001-5486-2093
Piero Baraldi https://orcid.org/0000-0003-4232-4161
Enrico Zio https://orcid.org/0000-0002-7108-637X

References

1. Ma X, Lan H, Qiao W, et al. On the causation correlation of maritime accidents based on data mining techniques. *Proc IMechE Part O: J Risk and Reliability* 2022; 0(0). DOI: 10.1177/1748006X221131717
2. Stewart M, Hodkiewicz M, Liu W, et al. MWO2KG and Echidna: constructing and exploring knowledge graphs from maintenance data. *Proc IMechE Part O: J Risk and Reliability* 2022; 0(0). DOI:10.1177/1748006X221131128
3. Wu H, French T, Liu W, et al. Automatic semantic knowledge extraction from electronic forms. *Proc IMechE Part O: J Risk and Reliability* 2022; 0(0). DOI: 10.1177/1748006X221098272
4. Ramos PMS, Macedo JB, Maior CBS, et al. Combining BERT with numerical variables to classify injury leave based on accident description. *Proc IMechE Part O: J Risk and Reliability* 2022; 0(0). DOI: 10.1177/1748006X221140194
5. Valcamonico D, Baraldi P, Amigoni F, et al. A framework based on Natural Language Processing and Machine Learning for the classification of the severity of road accidents from reports. *Proc IMechE Part O: J Risk and Reliability* 2022; 0(0). DOI: 10.1177/1748006X221140196