Reliability Engineering and System Safety 224 (2022) 108563

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

journal homepage: www.elsevier.com/locate/ress

Quantitative analysis of freight train derailment severity with structured and unstructured data

Bing Song [a,b], Zhipeng Zhang [a,*], Yong Qin [c], Xiang Liu [d], Hao Hu [b]

[a] State Key Laboratory of Ocean Engineering, Shanghai Jiao Tong University, China
[b] Department of Transportation Engineering, School of Naval Architecture, Ocean and Civil Engineering, Shanghai Jiao Tong University, China
[c] State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, China
[d] Department of Civil and Environmental Engineering, Rutgers, The State University of New Jersey, United States

ARTICLE INFO

Keywords:
Railroad safety
Derailment severity
Natural language processing
Topic modeling

ABSTRACT

Train safety has been a top priority in the railroad industry. Understanding accident risks is of paramount importance for prioritizing effective prevention strategies. Previous work has focused on estimating the severity of derailments and various statistical models based on structured data were used. However, unstructured data records which provide considerable information about train derailments have received minimal consideration due to a lack of procedures of processing and interpreting them. To narrow this knowledge gap, this study aims to quantitatively estimate derailment severity by considering unstructured data utilizing topic modeling. A statistical model that integrates both structured and unstructured data was established to analyze U.S. freight train derailments from 1996 to 2019. The comparative results of predictions revealed that the model with combined text information outperformed the one without the unstructured data.Quantile regression was also developed to assess various statistical distributions of derailment severity. Both models with unstructured data provide a deeper understanding of derailment severity and ultimately improve railroad safety performance.

1. Introduction

Railroads play a vital role in a nation's economy. Along with the vast economic benefits derived from rail transportation, safety is critical for the railroad industry and raises concerns from the public as well administrations [1]. Train accidents can lead to casualties, damaged infrastructure, interruption of services, and damage to the environment [2,3]. Based on the Federal Railroad Administration (FRA) records of train accidents in the United States (U.S.), train derailments have the highest frequency of occurrence among all types of train accidents [4]. Previous studies have primarily focused on the evaluation and quantification of railroad derailment frequency [5], particularly cause-specific derailment frequency studies. Moreover, it is also vital to explore and assess the magnitude and variability of train derailment severity [6,7].

Various factors may influence derailment severity, such as derailment speed, number of cars after the point of derailment, distribution of train power, proportion of loaded cars, derailment cause, and ground friction [6,8]. There is a continuous interest for government and railroad companies in understanding the effects of these factors on derailment severity, as these assessments can contribute to improving railroad safety by mitigating the consequences of train derailment.

Previous studies on this subject were performed based on derailment accident data using statistical modeling [9–12]. However, most researchers only considered structured data in the recorded derailment accident database owing to lack of techniques for processing unstructured data (e.g., text narratives) in accident records. To narrow the knowledge gap and improve train accident risk mitigation, this research aims to explore the utilization of topic modeling to leverage the value of unstructured data fields with estimation models of derailment severity and to improve the performance of railroad derailment severity prediction. A zero-truncated negative binomial (ZTNB) model that integrates both structured and unstructured data was developed, in which the coefficients of different extracted topics also provide additional insights inotreducing derailment severity.

Moreover, most of the earlier studies focused on developing various derailment models that concentrate only on the central tendency of the outcomes, while this study also pays attention to investigating how to use text analysis to deepen and strengthen the understanding of railroad derailment consequences with additional distributional statistics (e.g., quantiles). Thus, the quantile regression (QR) model integrating text

* Corresponding author.
E-mail address: zp.zhang@sjtu.edu.cn (Z. Zhang).

https://doi.org/10.1016/j.ress.2022.108563
Received 4 December 2021; Received in revised form 26 April 2022; Accepted 30 April 2022
Available online 2 May 2022
0951-8320/

information was also established in this study. The statistical models along with 24-year train derailment records show not only that the information extracted from unstructured derailment narratives can improve the performance of derailment severity estimation models, but also that it can strengthen quantitative understandings of derailment consequences. The ultimate objective of this study is to enhance the understanding of the railroad safety insights via text mining; therefore, effective accident prevention strategies can be adapted to reduce railroad transportation system risks. The methodology proposed in this study can also be adapted to analyze other railroad accidents, and ultimately to improve railroad safety and save lives.

2. Literature review

2.1. Train derailment severity

Various metrics have been used to evaluate the severity of train derailments in previous studies, such as the number of derailed cars [1,4], monetary damage [1], and casualties [14,15]. According to previous studies, the number of derailed cars is a suitable and trendy metric for evaluating the severity of train derailments. The generic term "cars" indicates all vehicles, including railcars, locomotives, and cabooses [6].

With targeted measures in the evaluation of derailment severity, simulations and statistical analysis are two basic approaches used in previous studies to model train derailment severity. The basis of simulation models in train derailment severity is generally related to nonlinear wheel–rail interaction. The reaction between the railroad vehicles and specific environment-related and track-related conditions can be visualized using physics law-driven models [6,9]. In terms of statistical derailment models, historical derailment accident data were primarily utilized to estimate the severity of train derailment. Compared with simulation models, statistical models rely significantly more on historical data. Consequently, statistical models have attracted the attention of academia owing to the rapid development of data processing technologies. Barkan et al. [1] explored the influence of derailment speed based on modeling the relationship between derailment severity and derailment speed. To further improve the forecast accuracy, Bagheri [16] presented an improved model to estimate the probability of derailment by position. The distribution of the train power and proportion of loaded cars, as novel factors, were considered by Liu et al. [6]. In recent years, there have also beenseveral studies that have attempted to use statistical learning methods (both machine learning and deep learning) to study train derailment. Dindar et al. [17–18] utilized a Bayesian network to identify and classify the human errors and climate errors causing derailments, respectively. Then,a novel DAG (Directed Acyclic Graph) was proposed to analyze the effects on the derailment of these two types of errors. Xu and Saleh [19] reviewed the application of machine learning for railroad reliability engineering. More novel and accurate insights can be provided by machine learning as compared to traditional methods.

In addition to structured data, unstructured data in accident records and databases, such as fixed field entries and narratives, can also provide a better understanding of the factors contributing to accidents [20]. However, a majority of the above models, even statistical learning methods, focus on utilizing structured data in derailment accident databases only. To date, few studies have included large-scale analyses of the narratives of derailments that can further improve railroad safety.

2.2. NLP in transportation safety

NLP, also known as computational linguistics, is a rapidly developing field that processes both written and spoken languages using computer analysis. It is acknowledged to be an interdisciplinary field, utilizing concepts from linguistics, computer science, statistics, and machine learning or deep learning in general [21]. Spurred by the progress of distributed word representations, various NLP tasks (e.g., natural language understanding and machine translation) have been successful with deep learning architectures [22]. Global interest has grown in applying NLP to the comprehension and analysis of transportation accidents to improve transportation safetyandreliability. Previous studies have shown that different NLP tasks have been performed in the transportation area, including automatic record classification, topic modeling of accident records, identifying similar records, and some active learning tasks [23–24].

Even though most researchers do not focus on the railroad industry, previous studies have provided remarkable suggestions and solutions for enhancing railroad safety with NLP [25]. Kwon et al. [23] used two classification methods, namely the naïve Bayes classifier and decision tree classifier, to reveal the relative importance of the data fields with respect to the resulting severity level. Heidarysafa et al. [24] applied deep learning methods together with powerful word embeddings such as *Word2Vec* and *GloVe* to classify accident cause values for the primary cause field using the text in railroad accident narratives. Their work showed that text classification can be performed accurately when processing transportation accident records, which in turn can help improve safety performance.

Another vital application of NLP in transportation safety is information extraction and topic modeling. Pereira et al. [26] utilized a topic modeling technique to obtain exact information from ground traffic incident records in real time. This study found that the prediction model with text features presented errors up to 28% lower than those without text information. Brown [20] explored the combination of text mining and machine learning algorithms to automatically discover accident characteristics that can provide a better understanding of the contributors to severe train accidents.

2.3. Knowledge gap

Various studies have focused on estimating derailment severity using different statistical models. However, nearly all of these studies were established based on structured data, and only numerical data from accident databases were employed. To narrow the knowledge gap and leverage the values of unstructured data fields in railroad safety, this study aims to investigate the derailment consequences that are defined as the number of derailed cars in the consist, using both numerical accident data information and unstructured accident narratives. The derailment severity statistical models (e.g., ZTNB model and QR model), along with improved severity prediction performance, can further deepen the understanding of derailments, and ultimately strengthen railroad safety based on better developed and allocated accident prevention strategies.

3. Topic modeling

3.1. LDA

LDA is a probabilistic document topic generation model and is a useful approach to uncover the latent topic structure and respective keywords for a corpus D [27]. This method can also be used to identify hidden subject information in a large-scale document collection. It contains a three-layer structure of words, topics, and documents, as shown in Fig. 1. The LDA model is a generative process for documents composed of topics with words. The general process of LDA is defined as follows [28]:

(1) For each topic, $k = \{1, \ldots, K\}$.
   (a) Draw a distribution over the vocabulary $V, \beta_k \sim Dir(\eta)$.
(2) For every document d.
   (a) Draw a distribution over topics, $\theta_d \sim Dir(\alpha)$(i.e., per-document topic proportion).
   (b) For each word w within document d.

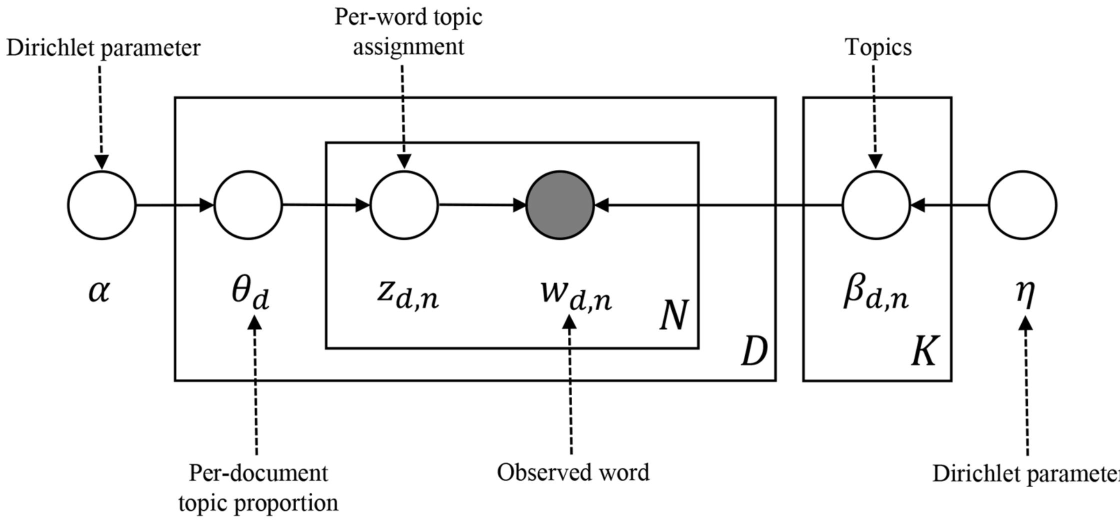

Fig. 1. Concept of LDA Algorithm.

(i) draw a topic assignment, $z_{d,n} \sim Multi(\theta_d)$, where $z_{d,n} \in \{1, \ldots, K\}$(i.e., per-word topic assignment).
(ii) draw a word $w_{d,n} \sim Multi(\beta_{z_{d,n}})$, where $w_{d,n} \in \{1, \ldots, V\}$.

Where $\beta_k$is the word distribution for topic k, θ_d is the topic distribution for document d, $z_{d,n}$ is the topic for the n-th word in document d, and $w_{d,n}$ is the n-th word in document d.

Each topic β_k is a multinomial distribution over the vocabulary V and comes from a Dirichlet distribution $\beta_k \sim Dir(\eta)$. Furthermore, every document is represented as a distribution over K topics and comes from a Dirichlet distribution $\theta_d \sim Dir(\alpha)$. Additionally, α and$\eta$are Dirichlet distribution parameters. The smoothing of topics within documents is denoted by α and the smoothing of words within topics is denoted byη. The joint distribution of all the hidden variables,β_K(topics), θ_D(per-document topic proportion),z_D (word topic assignments), and observed variables w_D (words in documents), is expressed below:

$$Pr(\beta_K, \theta_D, z_D, w_D | \alpha, \eta) = \prod_{k=1}^{K} Pr(\beta_k | \eta) \prod_{d=1}^{D} Pr(\theta_d | \alpha) \prod_{n=1}^{N} Pr(z_{d,n} | \theta_d) Pr(w_{d,n} | z_{d,n}, \beta_{d,k}) \tag{1}$$

The LDA algorithm is a classic bag-of-words approach. If the document set is an imbalanced dataset, in which the proportion of each topic is not equal and certain types of topics may constitute the majority, the output of the traditional LDA model will be repetitive and meaningless. To overcome this limitation, the term frequency–inverse documents frequency (TF-IDF) algorithm is combined with the classical LDA model to achieve an acceptable outcome in this study.

3.2. TF-IDF

The theoretical basis for the combination function of the TF-IDF algorithm was first established by Robertson [29]. As an improved bag-of-words model, the basis for TF-IDF is that a query term occurring in many documents is not a good discriminator and should be given less weight than one that occurs in just a few documents. In this study, TF-IDF was combined with the traditional LDA algorithm to overcome the limitation of the classical LDA algorithm. Many terms that appear frequently in a large number of derailment accident records, such as “train” and “car,” are considered less important than some words which appear within rare records repeatedly, such as “wheel,” “track,” and “gallon”. With the TF-IDF algorithm, the dataset can be converted into a more reasonable representation of word vectors, which are the inputs for the LDA model. The TF-IDF equation is as follows:

$$F - IDF = \frac{n_t}{N} * log \frac{K}{K_n} \tag{2}$$

Here,n_t is the occurrence of term t within a document; N is the number of terms in the document; K is the total number of documents; and $K_n$is the number of documents that contain term t. In this study, a document indicates the narrative of one derailment in the database.

4. Data

4.1. Data source

Accident information and records were obtained from the REA database, provided by the U.S. FRA. Railroads operating in the United States should submit a detailed accident record to the FRA if the damage cost of the accident to infrastructure and rolling stock exceeds a specified monetary threshold [11]. Considering the proportion of primary accident records that occurred on different tracks, this study only used accident data related to the main tracks [30].

In the FRA REA database, each recordable accident contains over 141 variables recorded as either structured or unstructured data. Detailed information on recorded train accidents, such as the accident cause, number of derailed cars, total monetary damage, track class, train length, and train speed, are recorded in the structured data. Unstructured data include free text segments that indicate a combination of 15 narrative fields [20]. These text fields provide brief descriptions of the accidents, written by the railroad authorities. Each field occupies a maximum of 100 bytes; therefore, each record occupies less than 1,500 bytes. Class I railroad derailment accidents that occurred between 1996 and 2019 were used in this research as the primary data source.

In the collected dataset, factors that contribute to the occurrence of these historical train derailments are also assessed to present an overview of these freight train derailments. Rolling stock factors, track factors, signal and train control system factors, human factors, and miscellaneous factors (e.g., off-road environmental factors, circumference intrusion factors) are taken into account in this analysis. In order to have an in-depth evaluation, these five categories of factors have been classified into ADL (Arthur D. Little) groupings,which are widely used in train accident analysis [31]. The top ten factors measured by severity and frequency have been selected and plotted in Fig. 2. Different shapes of markers represent the factors' categories. For example, broken rails or welds, as a track factor-related factor with red circular markers, have the highest average severity.

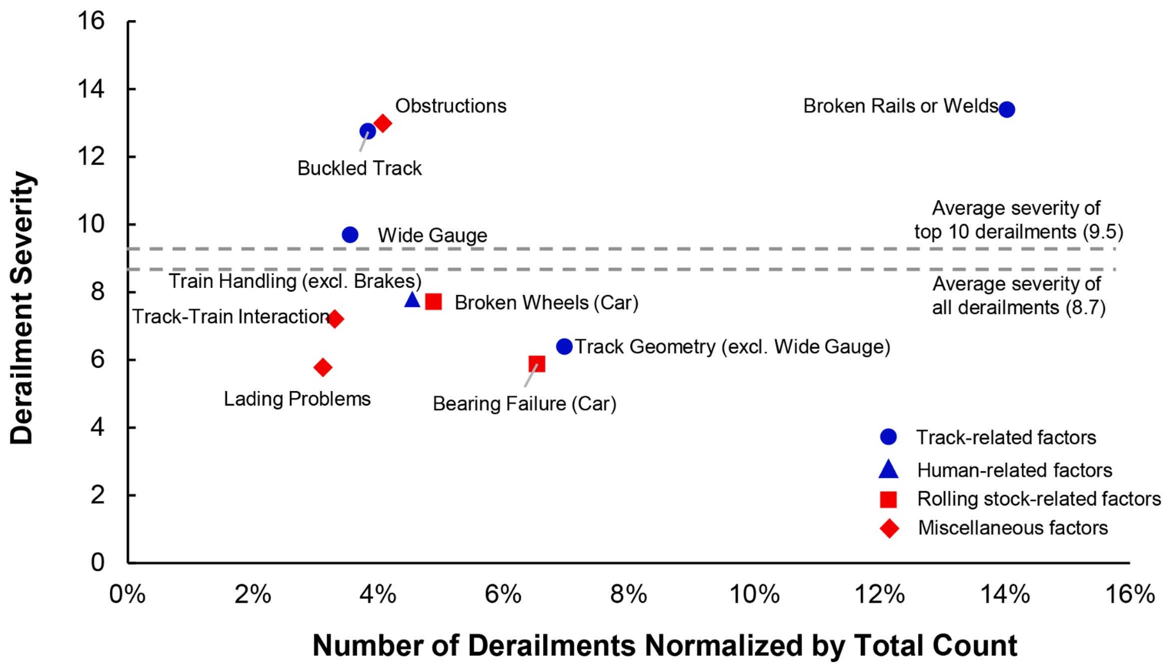

Fig. 2. Severity and frequency of the top ten derailment factors.

4.2. Explanatory variables from structured data

The objective of this study was to quantitatively investigate the impact of critical factors on derailment severity using estimation models by considering derailment narrative information. Law driven models (i.e., simulation models) and data-driven models (i.e., statistical analysis), as two main methods of derailment severity modeling, have revealed the factors which impact the consequences of train derailments [9]. Overall, train speed, train length and residual train length, train power distribution, proportion of loaded rail cars in the train, loading factor, coefficients of frictions, and peak coupler forces were expected to affect the number of derailed cars in one derailment [17–20]. Due to the data limitation, this paper is not able to incorporate all of these potential factors. Instead, four key potential influencing factors identified from previous studies [6,9] were included in the derailment severity estimation model, namely residual train length (RL), derailment speed (DS), train power distribution (DP), and proportion of loaded cars (LO) in the train. Narratives, as the unstructured data recorded for each train derailment, were also employed in the modeling and analysis. Overall, four structured factors and narratives take various aspects of train operations into account, such as line factors, rolling stock factors, off road environmental factors, and circumference intrusion factors. It is also acknowledged that alternative variables that were not considered in this study can be evaluated in future studiesusing the methodological framework proposed in this study.

4.2.1. Residual length

RL is the number of railcars following the point of derailment (POD), where POD is the position of the first car derailed. Previous researchers have found that RL has a complex relationship with the number of derailed cars when all other factors remain the same [16,31].

4.2.2. Derailment speed

DS is the train speed when a derailment accident occurs. Previous studies have shown that a higher DS results in more cars being derailed [8,16,28]. For example, Bagheri et al. [32] foundthat DS indicates the volume of kinetic energy during derailment, and a larger DS means more kinetic energy being produced, which results in a higher number of cars derailed.

4.2.3. Train power distribution

There are two common types of train power distribution, namely non-distributed power and distributed power (DP) [6]. In non-distributed power trains, only head locomotives are employed in freight trains and passenger trains. Instead, in DP trains, additional locomotives that are not head-end locomotives exist in other positions (typically in the middle and/or in the rear). In order to support the explanatory data analysis and distribution hypothesis examination, a binary variable is generated to represent these two types of freight trains. Specifically, 0 represents a non-distributed-power train and 1 represents a distributed-power train. Although DP is a binary variable in this study, mean values and standard deviations of DP are extracted in order to disclose additionalfeatures of their distributions. As shown in Table 1, the mean value of DP is 0.15, in which 1,160 derailed trains are distributed-power trains out of a total ofaround7,736 derailed trains. The standard deviation of DP (0.36) reflects that the distribution of DP in the studied freight trains is over-dispersive.

4.2.4. Proportion of loaded cars

LO is defined as the number of loaded cars normalized by the total number of cars (both empty and loaded) in the train [6]. Table 1a presents a summary and description of the above explanatory variables and Table 1b presents the Spearman correlation coefficients among these variables. Correlations between targets were also analyzed using the collected data (Table 1b). There is a significant correlation between DP and LO, in which the Spearman correlation coefficient is 0.278. This means that a derailed train with a higher LO usually has a higher likelihood of being equipped with DP. Due to the significant correlation between DP and LO, this study selected LO as the predictor variable and abandoned the variable of DP following previous research [6].

Table 1a
Description of variables for derailments of U.S. Class I Mainlines, 1996–2019. (a) Descriptive Statistics.

Variable	Mean	Stand Deviation	Minimum	Maximum	Type
RL	51.19	35.32	1	223	Count
DS	24.64	15.54	1	80	Continuous
DP	0.15	0.36	0	1	Binary
LO	0.70	0.33	0	1	Continuous

Table 1b
Description of variables for derailments of U.S. Class I Mainlines, 1996–2019. (b) Spearman correlation coefficients.

	DS	DP	LO
RL	0.007	0.09*	−0.04
DS	——	0.07*	0.007
DP	——	——	0.28**

* Spearman correlation coefficient between 0.05 and 0.20.
** Spearman correlation coefficient greater than 0.20.

4.3. *Preprocessing of unstructured data*

4.3.1. *Text cleaning*

Narratives in the REA database usually contain numbers and punctuation marks that are not useful during the LDA model training stage. Numerical elements, such as years, code of the train, and accident date need to be removed from the text data because even though they commonly possess a high occurrence frequency, they contain very little information. Furthermore, punctuation, such as commas, colons, semicolons, full stops, and question marks, should be deleted from the text to avoid generalization of the data. Negation is another issue that needs to be addressed in this step. For the LDA model, as a classical bag-of-words approach, it is vital to solve the negation problem because negation can affect the frequency of some words and ultimately influence the results of topic modeling. In this study, simple concatenation was used to solve this problem by generating relevant words. For example, “no injures” becomes “noinjures” after this step of preprocessing.

4.3.2. *Stop words removal*

Stop words are generally defined as the most frequent and common words in a language. Stop words of English can be divided into three categories, namely pronouns (e.g., I, you, we, they, he, she, and it), articles (a, an, and the), and other stop words (e.g., am, is, are, were, was, be, been, being, have, has, had, having, do, does, and doing). Stop words have a high occurrence frequency but low semantic relevance in the FRA REA database. Thus, it is essential to remove all English stop words in this step to ensure the accuracy of topic modeling.

4.3.3. *Tokenization*

There are three different concepts in NLP, namely token, type, and term. Token is an instance of a character sequence in some documents grouped as useful semantic units for processing. Type is the class of all tokens containing the same character sequence. Term is a normalized type included in the information retrieval (IR) system’s dictionary. Each narrative needs to be tokenized to obtain a unique word, and repeated words are only recorded once. Tokenization is a vital preprocessing procedure. In this study, the tokenization algorithm was provided by Natural language toolkit (NLTK). A simple example of tokenization is presented in the following. The original sentence is “the employee was standing by the switch.” After tokenization, this sentence will be converted into a series of single words (tokens): “the”, “employee”, “was”, “standing”, “by”, “the”, and “switch”.

4.3.4. *Stemming*

Stemming is the process of reducing each word to its word family root.The objective of stemming is to standardize different variations and inflections of words. The output of stemming is not necessarily a word that normally exists in English (e.g., “fly”and “flies” can be standardized into “fli”). For example, “seriously derailed” and “serious derailment” are both converted into “serious derail.”After the above steps, a dictionary to be used in topic modeling was prepared. Table 2 presents an example to demonstrate the effects of preprocessing. After the preprocessing, the number, stop words, and punctuation marks in the raw narrative are removed. Every letter is turned to its lowercase and the whole sentence is divided into tokens that are standardized by stemming.

Table 2
Effect of preprocessing for narrative

Raw narrative	Narrative after preprocessing
CREW WAS SWITCHING TRACK 8 AND CAR WAS KICKED DOWN THE LEAD TOWARD TRACK 8. EMPLOYEE WAS STANDING BY THE #8 SWITCH. CAR TIED ONTO A CAR ALREADY ON THE TRACK, CAUSING CONTENTS (SODIUM HYDROXIDE) TO SPLASH OUT OF CAR ONTO EMPLOYEE’’S CHEST AND ARM	‘crew’, ‘switch’, ‘track’, ‘car’, ‘kick’, ‘lead’, ‘track’, ‘employe’, ‘stand’, ‘switch’, ’car’, ‘ti’, ‘car’, ‘track’, ‘caus’, ‘content’, ‘sodium’, ‘hydroxid’, ‘splash’, ‘car’, ‘employe’, ‘chest’, ‘arm’

4.3.5. *Dictionary construction*

The purpose of the above steps was to reduce a wide set of words into a smaller set of relevant words that would be catalogued in a dictionary. Each word in this dictionary would therefore correspond to an index in the bag-of-words representation. Using this dictionary, the LDA model results could be obtained easily.

5. ZTNB model

Diverse count data regression methods have been used in accident analysis, according to previous research [33]. Among these, Poisson regression and negative binomial (NB) regression are two commonly used methods in accident analysis [34–35]. The Poisson model is suitable for data where the mean is equal to the variance, and the NB model assumes that the Poisson mean follows a gamma distribution. Some previous studies also used the NB model to analyze data whose variance was greater than the mean. Therefore, the NB model was more appropriate for estimating derailment severity in this study. However, because both the Poisson and NB distributions include zeros, they cannot be precisely used to analyze variables excluding zeros, such as the number of derailed cars. Koenker et al. [39] discussed methodologies for analyzing zero-truncated count data. Bayes’ theorem was used to calculate the probability of the response variable based on positive count data, which modified the previous research. The probability mass function (Eq. (3)), mean (Eq. (4)), variance (Eq. (5)), and likelihood function (Eq. (6)) of the ZTNB model are shown below. More detailed discussions of the ZTNB model can be found [36].

$$Pr(y_i|y_i>0)=\frac{\Gamma\left(y_i+\frac{1}{\alpha}\right)}{y_i!\Gamma\left(\frac{1}{\alpha}\right)\left[1-\left(\frac{1}{1+\alpha\mu_i}\right)^{\frac{1}{\alpha}}\right]}\left(\frac{\alpha\mu_i}{1+\alpha\mu_i}\right)^{y_i}\left(\frac{1}{1+\alpha\mu_i}\right)^{\frac{1}{\alpha}} \tag{3}$$

$$E(y_i|y_i>0)=\frac{\mu_i}{Pr(y_i>0)}=\frac{\mu_i}{1-\left(\frac{1}{1+\alpha\mu_i}\right)^{\frac{1}{\alpha}}} \tag{4}$$

$$Var(y_i|y_i>0)=\frac{E(y_i|y_i>0)}{\Pr(y_i>0)^{\alpha}}\left[1-\Pr(y_i=0)^{1+\alpha}\,E(y_i|y_i>0)\right] \tag{5}$$

$$L=\prod_{i=1}^{N}\Pr(y_i|y_i>0)=\prod_{i=1}^{N}\frac{\Gamma\left(y_i+\frac{1}{\alpha}\right)}{y_i!\Gamma\left(\frac{1}{\alpha}\right)\left[1-\left(\frac{1}{1+\alpha\mu_i}\right)^{\frac{1}{\alpha}}\right]}\left(\frac{\alpha\mu_i}{1+\alpha\mu_i}\right)^{y_i}\left(\frac{1}{1+\alpha\mu_i}\right)^{\frac{1}{\alpha}} \tag{6}$$

where α is the over-dispersion parameter; μ_i is the predicted derailment severity (number of derailed cars) for the ith observation; and y_i is the observed derailment severity for the ith observation. Then the response surface of the ZTNB model is given as

$$log(\mu_k)=\beta_0+\beta_1X_{1k}+\ldots+\beta_iX_{ik} \tag{7}$$

whereβ_k is the parameter coefficient of the kth predictor variable ($k=$

0 for the intercept) and X_{ik} is the value of the kth predictor variable for the ith observation. In this study, the RL, DS, LO, and topics obtained through LDA were utilized as the predictor variables.

The response variable in this study was the number of derailed cars. The dataset used was comprised of 7,736 freight train derailment records on Class I mainlines during the period 1996–2019, collected from the FRA REA database. This dataset was split into two datasets:a training set and a test set.The training set is made up of 75% of the collected data, which was used to fit the ZTNB model. The remaining 25% of the collected data formed the test set, which was used to examine the estimation performance of the fitted model. The established model considered the main effect, higher-order components, and interaction terms of the explanatory variables.

5.1. ZTNB Model without accident narrative information

To explore how narrative information can contribute to an enhanced derailment severity estimation model, a ZTNB model without the accident narrative information needed to be fitted as a benchmark for the modified model. Accordingly, this study first developed a ZTNB regression model without text information and chose variables within a 95% confidence interval (Table 3).

Consequently, the ZTNB model of train derailment severity with structured variables updated only with variable selection is presented below:

$$Z = \exp\left[0.525 + 0.016RL + 0.006DS - 0.083LO - 0.0001(RL)^2 + 0.78(LO)^2 + 0.0002RL \times DS + 0.01DS \times LO\right] \tag{8}$$

where$Z$ is the predicted number of derailed cars in train derailments on Class I mainlines.

5.2. ZTNB Model with accident narrative information

Based on the pre-processed unstructured accident narratives, as explained in Section 4.3, this section establishes a ZTNB model considering the accident narrative information. A comparison between the previous model and the modified model was also performed to interpret the contribution of the narrative information to estimating and understanding the consequences of train derailments.

5.2.1. LDA training and topic assignment

As mentioned in Section 3, there were three hyperparameters that needed to be defined and assigned before the training of the LDA model, namely K, α, and η. The latter two were vectors of dimensions K and W, respectively. The general practice commonly presumes that each element has the same value, often $1/K$ for both elements. The hyperparameter K represents the number of different topics that are expected to be obtained. In this study, to draw the top 10 topics from the derailment narratives, the LDA model with $K = 10$, $\alpha = 1/10$, and $\eta = 1/10$ was developed to assess the contributing roles in the derailment consequences. Table 4 presents a list of all 10 topics obtained from the historical derailment accident data with the developed LDA model.

Table 3
ZTNB model results for derailments of Class I mainlines, 1996–2019.

Variable	Coefficient	Standard error	*P*-Value
Intercept	0.525	0.090	$<10^{-3}$
RL	0.016	0.002	$<10^{-3}$
DS	0.006	0.003	$<10^{-3}$
LO	0.083	0.210	0.020
$(RL)^2$	1×10^{-4}	1×10^{-5}	$<10^{-3}$
$(LO)^2$	0.78	0.172	$<10^{-3}$
RL*DS	2×10^{-4}	3×10^{-5}	$<10^{-3}$
DS*LO	0.010	3×10^{-3}	$<10^{-3}$

Table 4
Ten topics obtained from train derailments from 1996 to 2019 with LDA.

Topic #0: 0.013*go + 0.013*caus + 0.012*travel + 0.012*main + 0.012*emerg + 0.011*journal + 0.011*track + 0.011*wind + 0.010*rail + 0.010*eastbound
Topic #1: 0.014*switch + 0.014*travel + 0.013*main + 0.013*break + 0.013*engine + 0.012*caus + 0.012*track + 0.012*emerg + 0.012*go
Topic #2: 0.015*track + 0.015*rail + 0.014*break + 0.013*main + 0.012*wheel + 0.011*damag + 0.011*caus + 0.010*unit + 0.010*stop + 0.010*inspect
Topic #3: 0.016*switch + 0.014*stop + 0.014*bnsf + 0.013*track + 0.012*crew + 0.012*travel + 0.011*emerg + 0.011*go + 0.011*drag + 0.011*inspect
Topic #4: 0.011*crew + 0.011*rear +0.011*caus + 0.010*rail + 0.010*curv + 0.010*track + 0.010*switch + 0.010*main + 0.010*engine + 0.010*load
Topic #5: 0.022*load + 0.019*unit + 0.019*empti + 0.018*ton + 0.016*head + 0.015*break + 0.015*journal + 0.014*rail + 0.013*caus + 0.013*south
Topic #6: 0.016*main + 0.014*emerg + 0.014*break + 0.014*go + 0.013*track + 0.013*rail + 0.013*cross + 0.012*line + 0.011*travel + 0.011*gallon
Topic #7: 0.018*east + 0.017*main + 0.017*travel + 0.016*head + 0.015*rail + 0.014*break + 0.014*track + 0.012*load + 0.011*pull + 0.011*emerg
Topic #8: 0.017*wheel + 0.013*caus + 0.012*tread + 0.011*truck + 0.011*switch + 0.011*clearanc + 0.011*main + 0.010*track + 0.010*build + 0.010*axl
Topic #9: 0.029*railcar + 0.025*track + 0.022*materi + 0.021*hazard + 0.021*release + 0.020*travers + 0.019*break + 0.017*singl + 0.017*main + 0.013*rail

5.2.2. Unique words and normalization

Once a topic assignment consequence was obtained, it would be a challenge to address the following two questions: (1) Do the narrative topics contain information that can be used as new variables in derailment severity estimation? and (2) How can the text be converted into a numerical feature as a new independent variable in the derailment prediction model. To address these questions, each topic, composed of unique words, should first be cleaned and normalized [16]. Table 5 shows the results for 10 topics after deleting repetitive words. From the pure words, preliminary insights can be collected as additional information on the topics. For example, Topic 6 is related to crossing accidents; Topic 8 concerns broken-rail or wheel-related derailments; and Topic 9 covers hazardous material leakage.

Second, the normalization procedure was used to convert extracted topics into numerical features by computing the proportion of topic words contained in the narratives. The range of normalization scores was from 0 to 1 [17]. For example, if an accident narrative m contained all the words in Topic n, the score S_{mn} for this accident and topic would be 1.0. If only 30% of the words of Topic n appeared in accident narrative m, S_{mn} would be equal to 0.3. Table 6 presents the descriptive statistics after the normalization procedure. As introduced in Section 3, LDA is an unsupervised learning algorithm, based on statistical principles and word frequency. Several topics obtained from the LDA need to be treated as irrelevant and need to be abandoned. However, instead of focusing on finding the best topic assignment of all the derailment narratives, the purpose of this study was to investigate how a narrative could improve the understanding of derailment severity. The topic

Table 5
Unique words in the 10 topics from derailment accident narratives.

Topic #0	Topic #1	Topic #2	Topic #3	Topic #4
Caus wind eastbound	Engin side	rail damag	switch stop inspect crew drag	rear curve
Topic #5	**Topic #6**	**Topic #7**	**Topic #8**	**Topic #9**
journal unit load empti ton south	Go emerg cross line gallon	travel main head east pull	wheel tread truck clearance build axl track break	materihazard releas travers singl

Table 6
Descriptive statistics of the normalization procedure.

	Topic #0	Topic #1	Topic #2	Topic #3	Topic #4	Topic #5	Topic #6	Topic #7	Topic #8	Topic #9
Mean	0.337	0.153	0.301	0.549	0.135	0.480	0.635	0.839	0.288	0.965
Std	0.513	0.377	0.502	0.819	0.360	0.857	0.768	0.912	0.584	1.261
Min	0	0	0	0	0	0	0	0	0	0
25%	0	0	0	0	0	0	0	0	0	0
50%	0	0	0	0	0	0	0	1.000	0	1.000
75%	1.000	0	1.000	1.000	0	1.000	1.000	1.000	1.000	1.000
Max	3.000	2.000	2.000	5.000	2.000	5.000	5.000	5.000	6.000	5.000

selection process is detailed in the development of estimation models.

5.2.3. ZTNB model with added narrative information

Following the aforementioned process, the first step was to choose the topics employed as new factors in the ZTNB model. The guideline was that only a topic whose *P*-value was less than 0.05 could be seen as a new variable. Ultimately, Topics 0, 2, 3, 6, and 8 were adopted as final model variables. The identical training set and test set used in Section 5.1 were also employed to develop the ZTNB regression model with narrative information and to examine the prediction performance. The final results of the modified ZTNB model are listed in Table 7.

Spearman correlation coefficients are calculated among Topics and Structured Data to validate that predictor variables are independent. The calculation results are presented in Fig. 3. Since all spearman correlation coefficients are less than 0.3, it is reasonable to conclude that topics and structured data are independent of each other.

The mean square error (MSE), root mean square error (RMSE), and mean absolute percentage error (MAPE) were selected as performance evaluation metrics. Eqs. (9)–(11) present how these metrics are calculated.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 \tag{9}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{10}$$

$$APE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \tag{11}$$

Where $\hat{y}_i = \{\hat{y}_1, \hat{y}_2, \ldots\ldots \hat{y}_n\}$represents a prediction value, and $y_i = \{y_1, y_2, \ldots\ldots y_n\}$ represents the observed values in derailment severity.

A comparison of the prediction accuracy between the two models is presented in Fig. 4, which reveals a clear contribution of narrative information to model performance. As shown in Fig. 4, adding narrative information to the ZTNB model can reduce the model's prediction errors. The MSE, MAPE, and RMSE decreased by 27.25%, 7.65%, and 12.81%, respectively. The value of MSE is considerably greater than MAPE and RMSE since MSE measures squared value of derailment severity, while MAPE and RMSE are the evaluations of order of derailed cars in one accident. Thus, the comparison demonstrates that a statistical model with narrative information could significantly improve derailment severity estimation and improve the performance of the built model. In addition, regardless of whether the text information is considered in the ZTNB models, the computation times of fitting two regression models have minimal dissimilarity.

Table 7
ZTNB model results for derailments with narrative information.

Variable	Coefficient	Standard error	*P*-value
Intercept	0.785	0.088	$<10^{-3}$
RL	0.018	0.002	$<10^{-3}$
DS	0.005	0.003	0.039
LO	–0.913	0.198	$<10^{-3}$
(RL)2	1×10^{-4}	1×10^{-5}	$<10^{-3}$
(LO)2	0.812	0.163	$<10^{-3}$
RL*DS	1×10^{-4}	3×10^{-5}	$<10^{-3}$
DS*LO	0.012	0.003	$<10^{-3}$
Rate 0	0.277	0.090	0.002
Rate 2	0.146	0.061	0.016
Rate 3	–0.686	0.096	$<10^{-3}$
Rate 6	0.491	0.099	$<10^{-3}$
Rate 8	–0.230	0.139	$<10^{-3}$

5.2.4. Discussion

The performance comparison of the ZTNB model with only structured data and the ZTNB model with topics revealed that narrative topics contained information that could be used as new variables to enhance the estimation of derailment consequences. It is vital to understand how the performance of the estimation model can be improved by integrating the narrative information and what additional insights about derailment severity can be provided.

Enhanced estimation performance with additional information from narratives: Previous researchers have concluded that derailment causes are an important factor in accident severity [4]. With the same physical conditions (e.g., derailment speed or loading proportion), different accident causes will lead to differences in the number of derailed cars. For instance, the average number of cars derailed per derailment caused by joint bar defects and track geometry (excluding wide gauge) were 15.8 and 6.5, respectively [4]. Therefore, there is a need to leverage the value of accident causes in the accident database to achieve an acceptable prediction accuracy for derailment severity.

Accident cause information was recorded in both structured and unstructured data in the REA database. In terms of the structured data, one field named "CAUSE" contained the primary accident cause code classified by the FRA. However, to the authors' knowledge, previous studies and practices have demonstrated a lack of consideration of additional accident causes extracted from accident narratives. In addition, several previous studies have attempted to utilize NLP to classify narratives according to the cause [6,17]. Topic modeling algorithms, such as LDA, can be utilized to extract keywords and assign topics from a document dataset. Once the keywords are converted into numerical features, additional information related to the cause of one train derailment can be absorbed by the prediction model and eventually help assess the derailment consequence with enhanced derailment severity estimation performance.

The FRA database records more than 300 accident-cause codes. For example, the keyword "switch" in Topic 3 can also be found in accidents with several FRA causecodes. If narratives contain the word "switch," the causes of derailments correspond to the above cause codes. When the text information is converted into numerical features, the estimation model will also be revised accordingly to obtain better accuracy. Narrative 1 gives an example containing "switch."

Narrative 1: Subgrade failed at mo 439.7 causing lead axle of the bnsf 8854 to derail axle, partially suspended by right hand shock and left/right chains. During eastward movvement, derailed axle struck the **switch** at mp 437.5 at which time, wheels were rerailed. Tie clips were intermittently destroyed between mp 439.7 and 437.5.

Moreover, not every term extracted by the LDA algorithm was directly related to the cause of the derailment. Even though some

Fig. 3. Spearman correlation coefficients matrix.

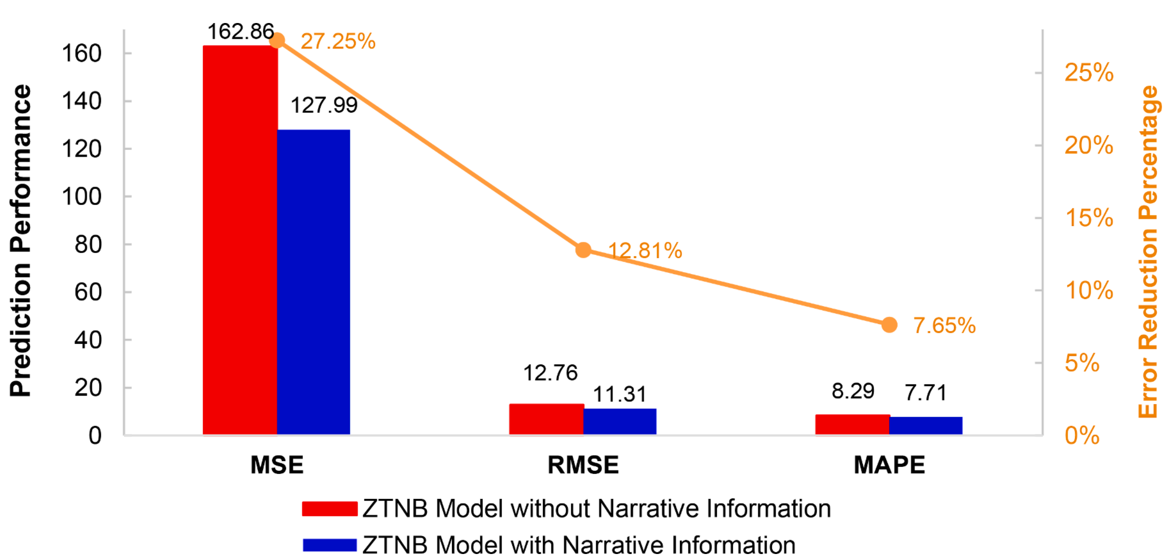

Fig. 4. Comparison of prediction accuracy between two models.

keywords could not be directly found in the FRA cause codes, these words still contained some important information associated with derailment severity to a certain extent. For instance, the occurrence of the keyword “gallon” from Topic 6 usually meant that the train involved in this derailment transported hazardous materials, in which case a large amount of leaked liquid cargo could have potentially more serious consequences. Overall, some keywords obtained from the LDA algorithm act like a banner or a sign symbolizing that this derailment may have more or less severe consequences. When this information is the additional input for the model, the estimation of derailment severity can be adjusted to achieve a slightly better performance.

Supplementary insights and understanding of derailments: The ultimate objective in exploring whether unstructured data (e.g., derailment narratives) can contribute to modeling derailment consequences was to utilize text mining of derailment narratives to gain an in-depth understanding of the severity of derailments and improve railroad safety. This section qualitatively describes how this study could help deepen and strengthen the understanding of derailments with accident narratives.

For example, there was one word “wind” in Topic 0. Clearly, if there was a strong wind or similar extreme weather at the time of derailment, the consequence of the derailment would typically be very severe. However, there were fewer than 100 derailments from 1996 to 2019 that contained the word “wind.” Its irregularity and low frequency made it very inconvenient to analyze this type of accident. With a text analysis of the derailment narratives, although the frequency of derailments that happened in extreme weather is very small, the occurrence of extreme-weather-related topics should receive more attention from railroad companies and administrations.The relationship between the insights provided by the narrative text and the understanding of the derailments were also highlighted by Brown [20]. In fact, railroads operating in areas with a high incidence of extreme weather conditions (e.g., extreme wind) should be more cautious and sensitive to extreme weather conditions. In addition, the discussion about the word “wind” also indicates that the LDA algorithm improved by TF-IDF could extract the events that had low frequency but severe consequence. Narrative 2 with the keyword “wind” is given below:

Narrative 2: ZSEME-22 went into emergency. The conductor found that the container had been blown off the car due to 60 plus mph **winds** causing ttex 353,239 to derail. 30. Method of operations: o. Other = RADIO.

Exclusions of several topics utilized in ZTNB regression model: Only five topics (topics 0, 2, 3, 6, and 8) were utilized in the final ZTNB model with narrative information based on two criteria. First, as an unsupervised learning approach, LDA is based on statistical principles and word frequency. The frequency of words in a narrative largely determines whether a word will be selected. Topics 1, 4, and 5 were obtained based on the LDA algorithm; however, they contradict people’s understanding of the topic distribution. In other words, these were some unsatisfactory results because the extracted words had a high frequency of appearance in derailment records, but do not contain important information. Even though the error caused by the LDA algorithm was almost impossible to eliminate, an increasing number of topics that could be used as new variables for the ZTNB model would be extracted.This would reduce the error of LDA and continue to improve the estimation model performance [17].

The second reason is associated with the evaluation metric of derailment severity. Specifically, Topic 9 denotes derailments with hazardous material. Evidently, a train carrying hazardous materials would cause more monetary loss. This study selected the number of derailed cars, instead of monetary loss, as an indicator of the severity of the derailment. A previous study disclosed that the hazardous materials such as petroleum, alcohol, or cheap ordinary goodswould not affect the number of derailed cars. However, in a study with monetary loss as an evaluation metric, Topic 9 may be a new factor in the estimation model.

5.3. Limitation of ZTNB model

The ZTNB model is appropriate for investigating datasets without zeros in the dependent variable. However, the ZTNB model can only analyze the mean response variable [6,11]; it still has several limitations as a classic count data model.The ZTNB model was used to predict the mean value, which means that it could not entirely represent the distribution of the data. Fig. 5 shows thedistribution of the cumulative percentages of the observed and prediction data. As shown in Fig. 5, the ZTNB model produced large errors at the beginning of the dataset. In addition, the mean value of the number of derailed cars was 8.673. If the prediction model solely considered the mean value, it might overestimate the derailment severity in accidents with only one or two cars derailed, or even underestimate the consequences of severe derailments [37]. To achieve additional statistical information and comprehensively consider other distributions, a QR model was established and the results were analyzed with the targeted dataset and objectives in this study.

6. Quantile model with unstructured data

To obtain additional statistical distributions of the number of cars derailed and comprehensively evaluate the derailment severity, a QR model considering derailment narrative information was developed, and will be discussed in the remainder of this section to gain more insights about derailment [38].

Mathematically, quantiles are obtained from the cumulative distribution function (CDF) of a random variable. For a random variable Y with a probability distribution function $F_Y(y)$,

$$F_Y(y) = Pr(Y \leq y) \tag{12}$$

Then the τth quantile of Y is defined as $Q_Y(\tau)$:

$$Q_Y(\tau) = F_Y^{-1}(\tau) = \ inf\ \{y : \ F_Y(y) \geq \tau\} \tag{13}$$

In particular, τ varies between 0 and 1, and the median is $Q(1/2)$. The QR model could provide a supplementary understanding of accident severity. More details about the QR model and how to establish the same can be found in Koenker et al. [39].

6.1. QR Model with accident narrative information

Owing to the limitations of the ZTNB model, it was essential to establish different QR models with the number of cars derailed as the dependent variable to obtain different quantiles of derailment severity. The intention of performing QR in this study was to explore more understanding provided by narrative information of derailment severity from another perspective and to support the establishment of the ZTNB model. Therefore, the dataset and selected independent variables in the QR model must be consistent with the factors in the ZTNB model to ensure that the following discussion is meaningful. Table 8 summarizes the QR estimates for the selected quantiles and the associated standard errors. As can be seen from the table, this study chose quantiles 0.2, 0.3, 0.4, 0.6, 0.8, and 0.9, along with the 0.5 quantile as the median value. Fig. 6 shows the variation trend of different independent variables with the growth of the quantiles.

6.2. Discussion on QR model-based analytical results

The statistical distributions of derailment severity with QR models revealed additional insights into the influence of narrative-based topics in different quantiles. It is also acknowledged that the robustness of the QR model resolved issues against outliers in the dataset. Even though a few outliers may have significantly affected the mean and variance, their effects on certain quantiles were lesser. Consequently, QR might still be practical in analyzing derailment severity using structured and unstructured data with a limited number of outliers.

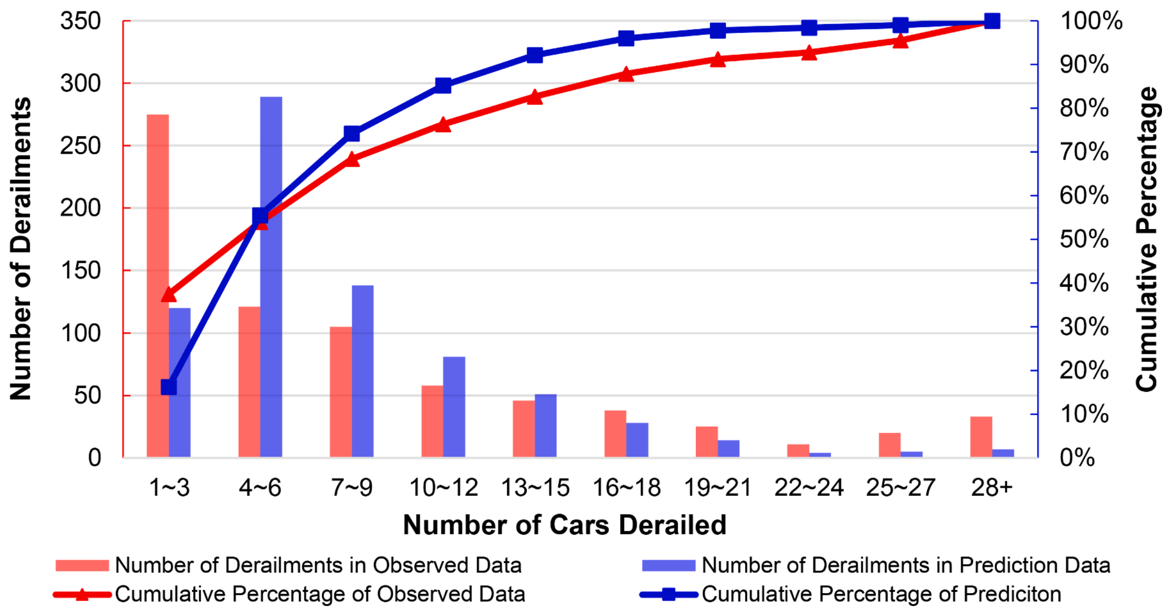

Fig. 5. Distribution of observed and prediction data.

Table 8
QR model results for derailments on Class I Mainlines, 1996–2019.

Variable	0.2 quantile Coeff*	P-value	0.3 quantile Coeff	P-value	0.4 quantile Coeff	P-value	Mean Coeff	P-value	0.6 quantile Coeff	P-value	0.8 quantile Coeff	P-value	0.9 quantile Coeff	P-value
Intercept	1	3×10^{-5}	2.298	0	3.222	0	3.933	0	4.229	0	5.601	0	6.424	0
RL	0	1	0.025	0	0.027	0	0.023	2×10^{-4}	0.031	0	0.058	0	0.091	0
DS	0	1	–0.005	0.511	–0.037	0	–0.068	0	-0.077	0	–0.049	0.018	–0.024	0.373
LO	0	1	–2.252	3×10^{-4}	–3.085	0	–2.783	1×10^{-4}	–2.423	0.004	–3.541	0.023	–1.932	0.314
$(RL)^2$	0	1	-2×10^{-4}	0	-3×10^{-4}	0	-3×10^{-4}	0	-5×10^{-4}	0	-7×10^{-4}	0	-9×10^{-4}	0
$(LO)^2$	0	1	2.856	0	3.327	0	2.707	0	2.189	0.002	2.739	0.022	1.096	0.490
RL*DS	0	1	2×10^{-4}	0.118	0.002	0	0.004	0	0.005	0	0.006	0	0.006	0
DS*LO	0	1	0.002	0.890	0.064	0	0.116	0	0.152	0	0.175	0	0.177	0
Rate 0	0	1	0.606	0.041	0.179	0.566	0.133	0.739	–0.045	0.927	0.813	0.222	1.775	0.041
Rate 2	0	1	0.598	0.008	1.034	0	1.183	0	0.988	0	0.310	0.435	0.716	0.267
Rate 3	0	1	–2.308	0	–3.312	0	–3.889	0	–3.769	0	–3.263	0	–1.557	0.077
Rate 6	0	1	2.830	0	3.193	0	2.913	0	2.761	0	2.744	0	1.873	0.043
Rate 8	0	1	–7.360	0	–10.758	0	-11.525	0	–11.228	0	–10.36	0	–10.746	0

* Coeff means coefficient.

The purpose of establishing the QR model was to investigate the different quantiles of derailment severity with narrative information and to support the establishment of the ZTNB model. More specifically, two preliminary inferences could be drawn from the signs and magnitudes of the coefficients to support the correctness of the ZTNB model. First, the explanatory variables kept the sign of the coefficients the same in different quantiles. Even though they may have different magnitudes in different quantile models, the same signs indicated that the influence of an explanatory variable on derailment severity was consistent. Second, the magnitudes of the parameter in the ZTNB model were between the 0.2 and 0.6 quantiles, which is reasonable because the mean was also a special quantile value. The two inferences based on the results of the QR models were similar to the conclusions from a previous study [6].

Moreover, several new insights were also obtained with the QR model-based analytical results that can potentially improve understanding of derailment severity. First, the *P*-values of several topics varied between different quantiles. This indicates that certain topics significantly affected the derailment severity only when it met a specific condition. For instance, the *P*-values of Topic 0 in 0.2, 0.3, 0.4, and 0.5 quantiles were greater than 0.1, which meant that Topic 0 played an insignificant role in the derailments (with less severe consequences). The *P*-value of Topic 0 decreased significantly in 0.8 and 0.9 quantiles.

A higher quantile indicates more severe derailment. Therefore, it can be speculated that derailments with narratives classified as Topic 0 may occur occasionally, but with serious consequences. As mentioned in Section 5.2.4, Topic 0 represents a derailment caused by extreme weather (e.g., wind). Even though derailments caused by extreme winds are not frequent, these derailments may result in very severe consequences in general. In addition, the magnitude of several topics appeared to be changing with an increase in the quantiles. For example, the magnitude of the coefficients in Topics 3 and 8 declined with an increase in the quantiles. This suggests that even though two topics have an influence on derailment severity, they could play a more vital role in derailments with less severity. Based on the ZTNB model and QR model, additional insights provided by derailment narratives about derailment severity can eventually help reduce the severity of derailment and improve the safety performance of railroad systems.

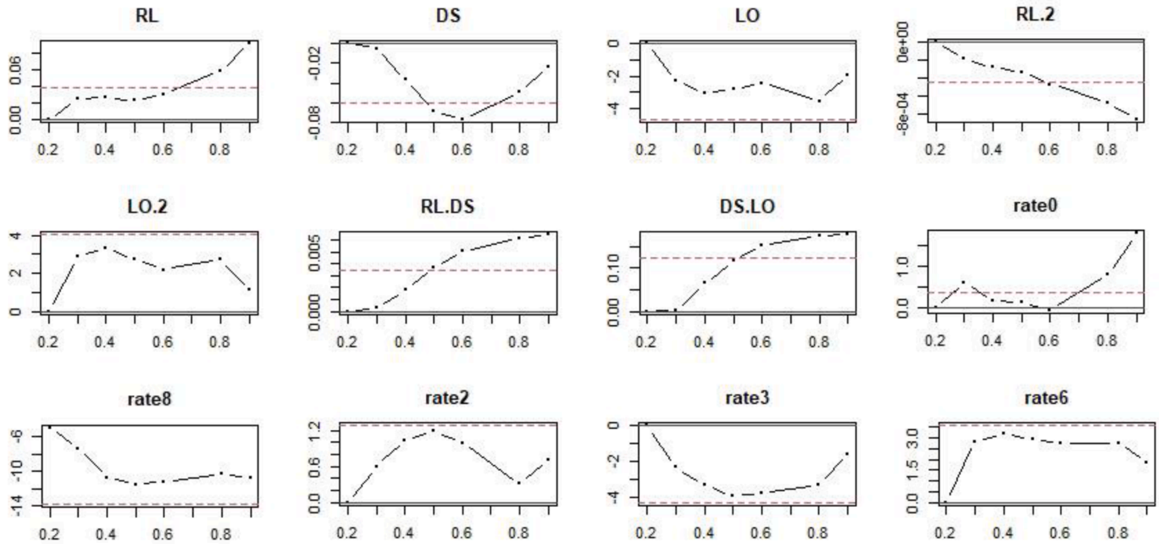

Fig. 6. Variation trend of independent variables with different quantiles.

7. Conclusions and future work

To explore derailment severity using narrative information, this study developed statistical modelsbased on historical train derailment accidents involving both structured data and unstructured data. Derailment severity was investigated with the ZTNB model with narrative information that was further compiled into the top five topics via Latent Dirichlet Allocation. The model yielded significant improvements (i.e., 27.25% in MSE, 7.65% in MAPE, and 12.81% in RMSE) in the accuracy of derailment severity estimation as compared with the model without narrative information. Additional distribution and quantile values of derailment severity were estimated using QR, involving derailment narrative information. The case-specific analysis showed that the model built with unstructured data could provide new insights on derailment severity, including extensive cause information, global features in the environments, and far-reaching characteristics of trainsThe combination of the narrative information in the FRA REA database and the statistical model of derailment severity can facilitate a better understanding of derailment severity distribution and influencing factors, thus ultimately contributing to more comprehensive strategies for train accident risk mitigation.

Several studies can be conducted in the future to further improve the assessment of train derailment severity. First, alternative approaches can be developed to convert narratives into more exhaustive numerical features. Deep learning algorithms (e.g., Bi-LSTM and Text-CNN) can be used to extract unstructured information. More advanced natural language processing algorithms should be involved to mitigate the errors caused by polysemy in future studies. Additionally, the interaction terms between topics and variables from structured data can be considered in the future and may provide more insights about the severity of derailment.

CRediT authorship contribution statement

Bing Song: Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft. **Zhipeng Zhang:** Conceptualization, Resources, Funding acquisition, Supervision, Writing – original draft, Project administration. **Yong Qin:** Validation, Investigation, Data curation. **Xiang Liu:** Methodology, Writing – review & editing. **Hao Hu:** Investigation, Project administration, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (52102413). The first author and second author were also sponsored by the Shanghai Sailing Program (21YF1420000) and the State Key Laboratory of Rail Traffic Control and Safety (RCS2021K010).

References

[1] Barkan CP, Dick CT, Anderson R. Railroad derailment factors affecting hazardous materials transportation risk. Transp Res Rec 2003;1825(1):64–74.
[2] Zhang Z, Turla T, Liu X. Analysis of human-factor-caused freight train accidents in the United States. J Transp Saf Secur 2019;0:1–29. Taylor & Francis.
[3] Boakye J, Guidotti R, Gardoni P, Murphy C. The role of transportation infrastructure on the impact of natural hazards on communities. Reliab Eng Syst Saf 2022;219:108184. https://doi.org/10.1016/j.ress.2021.108184.
[4] Liu X, Saat MR, Barkan CP. Analysis of causes of major train derailment and their effect on accident rates. Transp Res Rec 2012;2289(1):154–63.
[5] Anderson RT, Barkan CP. Railroad accident rates for use in transportation risk analysis. Transp Res Rec 2004;1863(1):88–98.
[6] Liu X, Saat MR, Qin X, Barkan CPL. Analysis of U.S. freight-train derailment severity using zero-truncated negative binomial regression and quantile regression. Accid Anal Prev 2013;59:87–93.
[7] Rungskunroch P, Jack A, Kaewunruen S. Benchmarking on railway safety performance using Bayesian inference, decision tree and petri-net techniques based on long-term accidental data sets. Reliab Eng Syst Saf 2021;213:107684.
[8] Anderson RT. Quantitative Analysis of Factors Affecting Railroad Accident Probability and Severity. University of Illinois at Urbana-Champaign; 2005.
[9] Martey EN, Attoh-Okine N. Analysis of train derailment severity using vine copula quantile regression modeling. Transp Res Part C Emerg Technol 2019;105: 485–503.
[10] Zhang Z, Liu X. Safety risk analysis of restricted-speed train accidents in the United States. J Risk Res 2020;23(9):1158–76.
[11] Sedghi M, Kauppila O, Bergquist B, Vanhatalo E, Kulahci M. A taxonomy of railway track maintenance planning and scheduling: a review and research trends. Reliab Eng Syst Saf 2021;215:107827.
[12] Bellè A, Zeng Z, Duval C, Sango M, Barros A. Modeling and vulnerability analysis of interdependent railway and power networks: application to British test systems. Reliab Eng Syst Saf 2022;217:108091.
[14] Liu X. Statistical temporal analysis of freight train derailment rates in the United States: 2000 to 2012. Transp Res Rec 2015;2476(1):119–25.
[15] Jeong D, Lyons M, Orringer O, Perlman A. Equations of motion for train derailment dynamics. In: Proceedings of the Rail Transportation Division Conference; 2007.
[16] Bagheri M. Risk-based model for effective marshalling of dangerous goods railway cars. UWSpace 2010. http://hdl.handle.net/10012/4996.

[17] Dindar S, Kaewunruen S, An M. Bayesian network-based human error reliability assessment of derailments. Reliab Eng Syst Saf 2020;197:106825.
[18] Dindar S, Kaewunruen S, An M. A hierarchical Bayesian-based model for hazard analysis of climate effect on failures of railway turnout components. Reliab Eng Syst Saf 2022;218:108130.
[19] Xu Z, Saleh JH. Machine learning for reliability engineering and safety applications: review of current status and future opportunities. Reliab Eng Syst Saf 2021;211:107530.
[20] Brown DE. Text mining the contributors to rail accidents. IEEE Trans Intell Transp Syst 2016;17:346–55.
[21] Cambria E, White B. Jumping NLP curves: a review of natural language processing research. IEEE Comput Intell Mag 2014;9(2):48–57.
[22] Sun S, Luo C, Chen J. A review of natural language processing techniques for opinion mining systems. Inf Fusion 2017;36:10–25.
[23] Kwon OH, Rhee W, Yoon Y. Application of classification algorithms for analysis of road safety risk factor dependencies. Accid Anal Prev 2015;75:1–15.
[24] Heidarysafa M, Kowsari K, Barnes L, Brown D. Analysis of railway accidents' narratives using deep learning. In: Proceedings of the 17th IEEE International Conference on Machine Learning and Applications (ICMLA); 2018.
[25] Ali F, El-Sappagh S, Kwak D. Fuzzy ontology and LSTM-based text mining: a transportation network monitoring system for assisting travel. Sensors (Switzerland) 2019;19(2):234.
[26] Pereira FC, Rodrigues F, Ben-Akiva M. Text analysis in incident duration prediction. Transp Res Part C Emerg Technol 2013;37:177–92.
[27] Krestel R, Fankhauser P, Nejdl W. Latent dirichlet allocation for tag recommendation. In: Proceedings of the Third ACM Conference on Recommender Systems; 2009.
[28] Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. J Mach Learn Res 2003;3:993–1022.
[29] Robertson S. Understanding inverse document frequency: on theoretical arguments for IDF. J Doc 2004;60(5):503–20.
[30] Federail Railroad Administration (FRA) (2021). Accident Data as Reported by Railroads. https://safetydata.fra.dot.gov/officeofsafety/default.aspx.
[31] Lin CY, Rapik Saat M, Barkan CP. Quantitative causal analysis of mainline passenger train accidents in the United States. Proc Inst Mech Eng Part F J Rail Rapid Transit 2020;234(8):869–84.
[32] Bagheri M, Saccomanno F, Chenouri S, Fu L. Reducing the threat of in-transit derailments involving dangerous goods through effective placement along the train consist. Accid Anal Prev 2011;43:613–20.
[33] Xie Q, Xiao J, Gardoni P, Hu K. Probabilistic analysis of building fire severity based on fire load density models. Fire Technol 2019;55(4):1349–75.
[34] Wood G. Confidence and prediction intervals for generalised linear accident models. Accid Anal Prev 2005;37(2):267–73.
[35] Lord D, Mannering F. The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. Transp Res Part A Policy Pract 2010;44(5):291–305.
[36] Zuur AF, Ieno EN, Walker NJ, Saveliev AA, Smith GM. Zero-truncated and zero-inflated models for count data. Mixed Effects Models and Extensions in Ecology with R. Springer; 2009. p. 261–93.
[37] Liu X, Saat MR, Barkan CP. Probability analysis of tank-car release incidents in railway hazardous materials transportation. J Hazard Mater 2014;276:442–51.
[38] Torossian L, Picheny V, Faivre R, Garivier A. A review on quantile regression for stochastic computer experiments. Reliab Eng Syst Saf 2020;201:106858.
[39] Koenker R, Chernozhukov V, He X, Peng L. Handbook of Quantile Regression. Handbook of Quantile Regression. CRC Press; 2017.