Fusing Data Extracted from Bridge Inspection Reports for Enhanced Data-Driven Bridge Deterioration Prediction: A Hybrid Data Fusion Method

Kaijian Liu, Ph.D., A.M.ASCE[1]; and Nora El-Gohary, Ph.D., A.M.ASCE[2]

Abstract: Data buried in textual bridge inspection reports offer great promise for enhanced data-driven bridge deterioration prediction. However, learning from these reports is challenging because they typically use multiple concept names to refer to the same entity and typically describe multiple instances of the same type of deficiency. Such multiple names and instances increase the dimensionality and the sparsity of the feature space, which would cause overfitting to a particular feature, undermine the generalizability of the machine learning models, and compromise the performance of the data-driven prediction. To address this challenge, this paper proposes a new hybrid data fusion method. It combines an unsupervised named entity normalization method and an entropy-based numerical data fusion method to fuse concept names and numerical data, respectively. The proposed normalization method uses an n-gram model to generate candidate canonical identifier names and utilizes corpus statistics and lexical patterns to fuse the multiple concept names into a candidate name that balances abstraction and detailedness. The proposed fusion method uses data discretization and information entropy to fuse the multiple deficiency measures (of the instances) into a single representation. The hybrid fusion method was validated in fusing data extracted from textual bridge inspection reports for supporting the prediction of future bridge condition ratings. Learning from the fused data, compared to learning from the unfused data, improved the accuracies of predicting the ratings of decks, superstructures, and substructures by 8.0%, 8.5%, and 7.9%, respectively. **DOI: 10.1061/(ASCE)CP.1943-5487.0000921.**

Author keywords: Data fusion; Named entity normalization; Bridges; Deterioration prediction; Data-driven approach; Machine learning.

Introduction

Textual bridge inspection reports offer great promise for enhanced data-driven bridge deterioration prediction (FHWA 2013, 2014; Liu and El-Gohary 2016, 2017b; Li and Harris 2019). They include a large amount of detailed data about the element-level deterioration conditions of bridges (e.g., deficiency types and quantities, and maintenance action and material types)—much richer than other sources of data that are typically used for deterioration prediction [e.g., National Bridge Inventory (NBI) data (FHWA 1995), which describe bridge conditions mainly by condition ratings]. However, although many prediction methods/models have been developed in current practice (e.g., AASHTO 2018; FHWA 2020) and in the existing literature (e.g., Huang 2010; Chang et al. 2018; Lu et al. 2019), they can only learn from structured data such as NBI data, traffic data, and weather data. But, due to the challenges in analyzing textual data, existing methods/models are not able to also exploit the wealth of the condition data buried in the reports for improved performance of bridge deterioration prediction.

In addressing these challenges, information extraction methods (e.g., Liu and El-Gohary 2016, 2017b) have been developed for extracting information about bridge conditions from the reports into a structured data format. However, directly learning from the extracted data is still a great challenge because different inspection reports typically use multiple concept names to refer to the same entity and typically describe multiple instances of the same type of deficiency. For example, even in the same report (WSDOT 2013), multiple concept names (e.g., "deep edge spall" and "deep top edge spall") are used to refer to the same deficiency entity (e.g., "spall") and this entity has two instances at different locations on the same bridge element entity (e.g., "bridge rail"), which is referred to in the report also using multiple names (e.g., "bridge rail" and "north concrete bridge rail"): "At SW corner, behind the thrie beam, bridge rail has an $18'' \times 6'' \times 3''$ deep edge spall" and "The north concrete bridge rail at the east end has a $3'' \times 6'' \times 3''$ deep top edge spall." Similarly, in the same report (NYSDOT 2015), multiple concept names (e.g., "deep spalled areas" and "deep spall") are used to refer to the same deficiency entity (e.g., "spall"), which has two instances at different locations on the same bridge element entity (e.g., "cap beam underside"): "The underside of cap beam exhibits up to $4' \times 3' \times 4''$ deep spalled areas . . . " and "The underside of the cap beam between columns C1 and C2 exhibits $3' \times 20'' \times 3''$ deep spall . . . " Such multiple names and instances are common in textual bridge inspection reports. They increase the dimensionality and the sparsity of the feature space, which would cause overfitting to a particular feature, undermine the generalizability of the machine learning models, and compromise the performance of the data-driven prediction.

There is, thus, a need to fuse data extracted from bridge inspection reports into a unified representation for supporting enhanced data-driven bridge deterioration prediction. Such fusion requires two tasks. First, concept names that refer to the same entity, but vary in terms of surface forms and abstraction levels, need to be fused into canonical identifier names. This is different from concept mapping (e.g., Zhang and El-Gohary 2016; Le and Jeong 2017),

[1]Assistant Professor, Dept. of Civil, Environmental and Ocean Engineering, Stevens Institute of Technology, 1 Castle Point Terrace, Hoboken, NJ 07030. Email: Kaijian.Liu@stevens.edu

[2]Associate Professor, Dept. of Civil and Environmental Engineering, Univ. of Illinois at Urbana–Champaign, 205 N. Mathews Ave., Urbana, IL 61801 (corresponding author). Email: gohary@illinois.edu

Note. This manuscript was submitted on January 31, 2020; approved on May 11, 2020; published online on August 29, 2020. Discussion period open until January 29, 2021; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Computing in Civil Engineering***, © ASCE, ISSN 0887-3801.

which focuses on classifying the types of relationships between concept names and mapping equivalent names together. Rather, this is a concept naming problem—representing the concept names using canonical identifier names that balance abstraction and detailedness, so that they are not too frequent or too rare (in a collection of reports) to the extent of causing the loss of distinctive feature patterns or undermining the generalizability. Fusing concept names was thus defined, in this research, as a named entity normalization task: the multiple concept names that are used in a single report to refer to the same entity are normalized into a canonical identifier with balanced abstraction and detailedness, and the identifiers from different reports are subsequently fused if they are the same. Second, the numerical measures of the multiple instances, which are of the same type of deficiency but are at different locations on a bridge element, need to be fused into a single representation. Unlike data in multisensor data fusion applications (e.g., Jiang et al. 2016; Zhang et al. 2017), which are mainly characterized as being conflicting, imprecise, and/or multimodal (Khaleghi et al. 2013), each of the deficiency measures is partially describing the overall condition of the deficiency, and these data are, thus, complementary. Thus, in this research, fusing deficiency measures was defined as a numerical data fusion task: the measures of the deficiency instances from one report are fused into a single representation that is representative of all the original measures.

To address these needs, the authors propose a new hybrid data fusion method. The proposed hybrid method combines an unsupervised named entity normalization method and an entropy-based numerical data fusion method to fuse concept names and numerical data, respectively. The proposed normalization method uses an n-gram model to generate candidate canonical identifier names from the original concept names and uses a concept ranking function and a selection rule, which consider both corpus statistics and lexical patterns of the names, to select an identifier that balances abstraction and detailedness. The proposed fusion method uses data discretization to define candidate representations for the multiple deficiency measures and leverages information entropy to quantify the representativeness of the candidates for selecting the candidate that is the most representative of the original measures.

Background

Named Entity Normalization

Named entity normalization transforms named entities (i.e., concept names) that refer to the same entity into a canonical identifier name (Liu et al. 2012). Existing normalization methods are dictionary-based or machine learning–based and mainly focus on dealing with the surface-form variations in concept names.

Dictionary-Based Named Entity Normalization

Dictionary-based methods rely on established lexicons in domain-specific dictionaries or domain-general knowledge bases (especially Wikipedia) to fuse concept names. The lexicons are used as a look-up source of identifier names. To find an identifier from the lexicons, corpus-based [e.g., pointwise mutual information (Church and Hanks 1990)] or knowledge-based [e.g., Jiang-Conrath similarity (Jiang and Conrath 1997)] concept similarity assessment methods are used to assess the similarity between a concept name and an identifier. In existing research efforts, domain-specific dictionaries have been utilized for fusing species and organism names (e.g., Pafilis et al. 2013), disease names (e.g., Wei et al. 2016), and biomedical names (e.g., Lee et al. 2016). Wikipedia has been used for supporting named entity normalization–related applications, such as text annotation (e.g., Mihalcea and Csomai 2007), knowledge base construction (e.g., Alhelbawy and Gaizauskas 2014), and question answering (e.g., Wang et al. 2017).

Machine Learning–Based Named Entity Normalization

Machine learning–based methods use machine learning algorithms to learn how to fuse concept names. A number of supervised algorithms have been used for developing normalization models, including support vector machines (e.g., Magdy et al. 2007), generalized perceptron (e.g., Wagner and Foster 2015), random forests (e.g., Jin 2015), conditional random fields (e.g., Akhtar et al. 2015), feedforward neural networks (e.g., Leeman-Munk et al. 2015), long short-term memory recurrent neural networks (e.g., Han et al. 2019), and Siamese recurrent neural networks (e.g., Fakhraei and Ambite 2018). Some of these models directly predict identifier concept names (e.g., Leeman-Munk et al. 2015), and some predict the edit operations (e.g., insert, replace, and delete) needed to convert concept names into their identifiers (e.g., Han et al. 2019). In either case, human-annotated data are required. Because of the challenges in annotating data, several unsupervised normalization methods have been developed (e.g., Yang and Eisenstein 2013; Liu and El-Gohary 2018; Tahmasebi et al. 2019). Although unsupervised methods do not require annotated data, they need a set of target identifiers as input in order to compute the similarities between concept names and identifiers (which makes them resemble dictionary-based methods).

Numerical Data Fusion

Numerical data fusion transforms numerical data (e.g., numerical deficiency measures)—either from a single source or different sources and/or at different time points—into a unified representation (Boström et al. 2007). Existing methods mainly use descriptive statistics or fusion theories to conduct data fusion.

Descriptive Statistics

Descriptive statistics quantitatively describe the features of a set of data (Mann 1995). The commonly used descriptive statistics in data fusion include the measures of data central tendency and the measures of data variation. Central tendency measures include arithmetic mean, Bonferroni mean, geometric mean, harmonic mean, Heronian mean, power mean, median, and mode. Variation measures include coefficient of variation, mean absolution deviation, range, standard deviation, and variance. For a detailed description of these measures, the readers are referred to Mendenhall and Sincich (2016). Although descriptive statistics are simple, they have been used in some data fusion applications and achieved certain levels of success. For example, using a set of descriptive statistics, Wimmer et al. (2008) fused audio and video features for emotion recognition, Zhang (2015) fused water-depth data and bathymetry data for creating benthic habitat maps, and Varga et al. (2018) fused pixel-level normalized difference vegetation indexes across time for land-cover analysis.

Data Fusion Theory

Several data fusion theories have been developed, including Dempster-Shafer theory (Shafer 1976), fuzzy set theory (Zadeh 1965), possibility theory (Zadeh 1978), and rough set theory (Pawlak 2012). Dempster-Shafer theory assigns a belief mass to a fused value (which could be a single number, interval, or set) based on the strength of the evidence supporting this value. In the presence of evidence from multiple sources, it uses a joint belief mass function to fuse the belief masses, where the function considers both the agreement and conflict levels of the evidence. It selects the fused value that has the largest belief mass to represent data from multiple sources. Fuzzy set theory is a theoretical reasoning scheme that uses partial set memberships of data to allow for imprecise, rather than crisp, reasoning (Khaleghi et al. 2013). The memberships of imprecise data to a fused value are quantified using a membership function (e.g., piecewise linear functions and Gaussian distribution function) and are then fused using an aggregation function (e.g., averaging, conjunctive, and disjunctive functions). The fused value that has the largest aggregated membership degree is used to represent imprecise data from multiple sources. Possibility theory, as an extension of fuzzy set theory, was developed to further deal with incomplete data using possibility and necessity measures, which quantify the plausibility and the certainty of a fused value given incomplete data, respectively (Destercke et al. 2008). Rough set theory could be applied to data fusion using lower and upper approximations to find a fused value that has the highest approximation accuracy for representing data from multiple sources. Despite being theoretically applicable, this theory has rarely been used in data fusion (Khaleghi et al. 2013).

State of the Art and Knowledge Gaps

A number of research efforts have been undertaken in the areas of named entity normalization and numerical data fusion. Despite the importance of these efforts, two primary knowledge gaps exist in each of the areas.

In the area of named entity normalization, there is a lack of methods that do not require human involvement in the normalization process. Most of the existing methods rely heavily on human-developed dictionaries or training data to normalize concept names (see the "Named Entity Normalization" section). However, despite the fact that several guidelines define the standard vocabularies used for structured bridge data (e.g., FHWA 1995; AASHTO 2010), there are no such guidelines for inspectors/writers—who have very different writing styles and specificity levels—to follow when choosing the concept names to use in the textual bridge inspection reports. As a result, the concept names used in the reports vary, to a high degree, in terms of surface forms and abstraction levels. It is challenging to develop normalization dictionaries/data that can representatively and comprehensively capture such high-degree variations. Second, there is a lack of normalization methods that are able to normalize concept names with both types of variations, such as those in bridge inspection reports. Most of the existing methods mainly focus on dealing with surface-form variations, which are caused by different naming conventions, e.g., acronyms and morphological variations. Yet they are limited in normalizing concept names that also vary in terms of abstraction levels (e.g., "north concrete bridge rail," a subconcept of "bridge railing"). Balancing the abstraction and detailedness of identifier names is critical to the machine learning–based bridge deterioration prediction model. As the features of the model, abstract identifiers (e.g., using "bridge" as the identifier of the aforementioned names) are too frequent in a collection of reports and, thus, lead to the loss of distinctive feature patterns. On the other hand, detailed identifiers (e.g., using "north concrete bridge rail") are too rare in the collection and, thus, increase the dimensionality and the sparsity of the feature space, which would cause overfitting to a particular feature and therefore undermine the generalizability of the model.

In the area of numerical data fusion, there is a lack of methods that define the interval-based representation of fused data in an objective way. Interval-based representations are usually used in major data fusion frameworks to characterize the uncertainty in the data (Sentz and Ferson 2002; Torra 2010). However, most of the existing methods (e.g., Zhang et al. 2017; Tian et al. 2018; He et al. 2018; Wu et al. 2018; Liu and El-Gohary 2019; Song et al. 2019) define the representation (i.e., defining the number of intervals and the size of each interval) in a subjective way. For example, based on subjective human judgment, Zhang et al. (2017) defined the representation of fused building settlement data as four equal-size intervals. Subjective judgments are limited in defining the optimal number of intervals and the optimal size of each interval, because there is a tradeoff between the two. A large number of intervals is preferred to capture more distinctive data instances for avoiding underfitting, and, at the same time, a large interval size is preferred to retain more data instances within an interval for avoiding overfitting. But as the number increases, the size decreases. Such a tradeoff is very difficult to balance using subjective judgments of humans. Second, there is a lack of fusion methods that focus on fusing complementary data, such as the numerical deficiency measures in inspection reports, each of which partially describes the overall condition of a deficiency. The majority of existing fusion methods (e.g., Zheng and Deng 2018; Xiao 2019; He et al. 2018; Mohammadi et al. 2019) focus on fusing data that are imprecise, conflicting, and/or multimodal (Khaleghi et al. 2013) using fuzzy set theory, Dempster-Shafer theory, and/or matrix factorization (Sentz and Ferson 2002; Lahat et al. 2015). When fusing complementary data, they would result in an interval-based representation that can only represent a subset of the data, which are less imprecise or conflicting but cannot fully capture the whole condition that all the data collectively describe. Thus, despite being successful in their intended applications, existing data fusion methods are limited in fusing complementary data.

Proposed Hybrid Data Fusion Method

A hybrid data fusion method is proposed. At the cornerstone of the method are two proposed submethods for fusing concept names and numerical data, respectively: an unsupervised named entity normalization method and an entropy-based numerical data fusion method. As depicted in Fig. 1, the input of the proposed hybrid method is data records that are extracted from different bridge inspection reports and are linked if they come from the same report and refer to the same entity. A data record is a structured

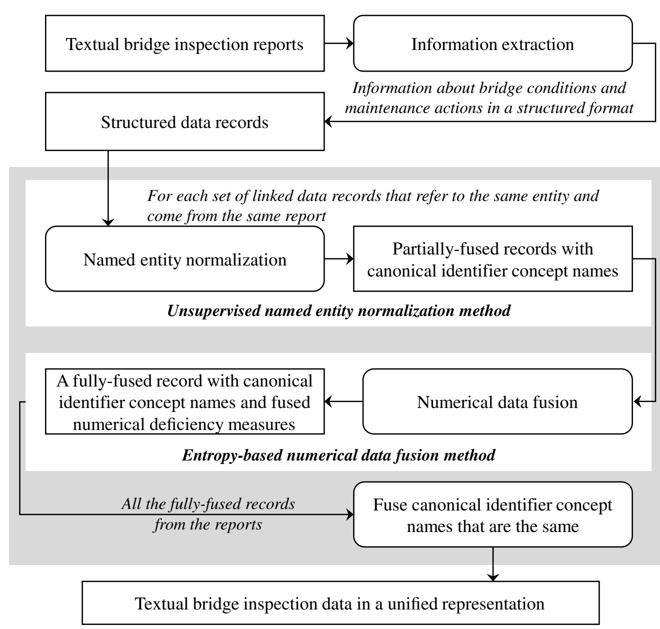


Fig. 1. Overview of proposed hybrid data fusion method.

representation of bridge condition and maintenance action information extracted from an inspection report. For example, <bridge element = "bridge rail," deficiency = "deep edge spall," deficiency length = "18," deficiency length unit = "ft"> is the (partial) data record of the information extracted from the following sentence: "At SW corner, behind the thrie beam, bridge rail has an $18'' \times 6'' \times 3''$ deep edge spall" (WSDOT 2013). The method includes two main steps for fusing these records. First, for a set of linked records, the concept names are fused into identifier names with balanced abstraction and detailedness using the proposed normalization method, resulting in a set of partially fused records. The identifier names from all the partially fused records are fused if they are the same, and the fused names are used as features in the unified representation of the reports. Second, for a set of partially fused records, the numerical deficiency measures are fused into a single interval-based representation using the proposed fusion method, resulting in a fully fused record. The fused data from all the fully fused records are used as values of their corresponding features/names and inspection reports in the unified representation.

Unsupervised Named Entity Normalization

A new unsupervised named entity normalization method is proposed. It fuses concept names that refer to the same entity, but vary in terms of both surface forms and abstraction levels, into a canonical identifier concept name that balances abstraction and detailedness. In bridge inspection reports, concept names are used to refer to entities about bridge elements, deficiencies, deficiency causes, maintenance actions, and maintenance materials. For example, in the reports by WSDOT (2013) and NYSDOT (2015), the following concept names are used to refer to the deficiency entity "spall": "deep edge spall," "deep top edge spall," "deep spalled areas," and "deep spall." The proposed method, as depicted in Fig. 2, includes three primary components: identifier concept name generation, ranking, and selection.

Identifier Concept Name Generation

Identifier concept name generation aims to generate all candidate identifier concept names—in their canonical forms and at different abstraction levels—that a set of original concept names could have. It includes two steps: morphological analysis and *n*-gram generation. Morphological analysis aims to analyze how a term is formed based on morphological derivation and inflection and to map the term into a canonical form. It was used to account for the surface-form variations. For example, for "bridge railing" and "bridge rail," morphological analysis removed the suffix "railing" and mapped the first name to its canonical form "bridge rail," resulting in a normalized surface form of the two. *N*-gram generation aims to generate candidate identifier names that are at different abstraction levels, so that an identifier name with balanced abstraction and detailedness can be subsequently selected. It was used to capture the abstraction-detailedness variations. Two types of candidate names are generated from the original names (in canonical forms) using an *n*-gram language model: regular and skip *n*-grams. Regular *n*-grams are the concept names (e.g., unigram, bigram, and trigram concept names) that have constituent terms following the same consecutive sequence as they appear in an original concept name. Skip *n*-grams are similar to regular *n*-grams, but their terms are not consecutive in the original name. For example, "asphalt deck" and "asphalt wearing" are the regular and skip bigrams of the concept name "asphalt deck wearing surface," respectively.

Identifier Concept Name Ranking

Identifier concept name ranking aims to rank the generated candidate identifier concept names. The ranking is conducted separately

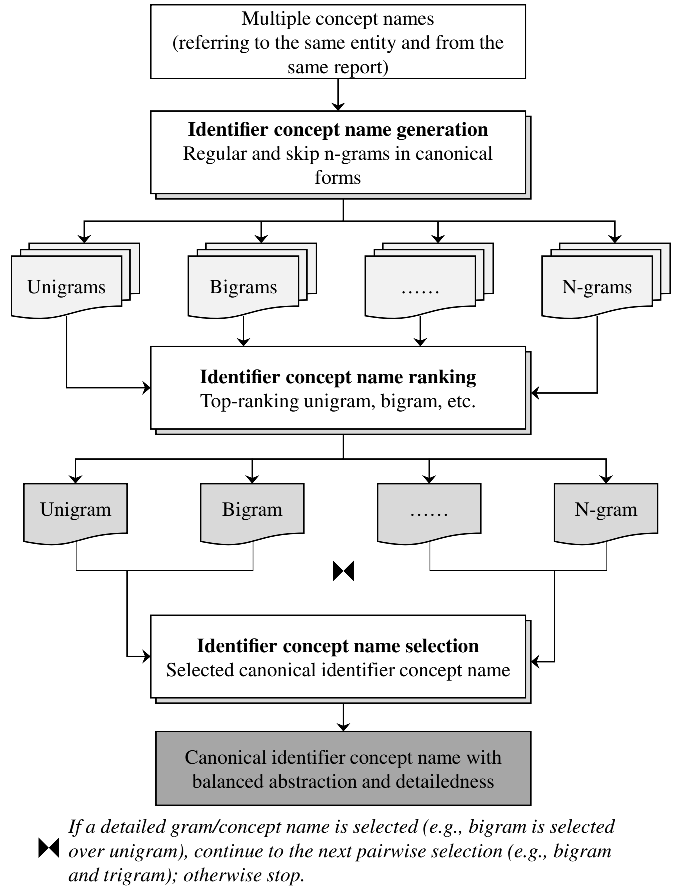


Fig. 2. Proposed unsupervised named entity normalization method.

at each abstraction level. For example, bigram names (both regular and skip) are ranked separately—not together with other types of names (e.g., unigram and trigram names)—to avoid the mixing of concept name distributions, which would negatively affect the ranking. A new concept ranking function is proposed to rank the identifiers. As shown in Eq. (1), the proposed function considers the corpus-statistic score (CSS), term-position score (TPS), and term-sequence score (TSS) of a candidate identifier concept name (CICN) to calculate its ranking score:

$$\text{Ranking score}(CICN) = CSS(CICN) \times TPS(CICN) \times TSS(CICN) \quad (1)$$

The CSS is used to rank the candidate names based on how frequent or rare they are in a collection of bridge inspection reports. To calculate the scores, two alternative corpus-statistic measures—term frequency (TF) and inverse document frequency (IDF)—were selected. TF captures the frequency rate of a concept name in all the sets of candidate names, where each set contains the names that refer to the same entity and come from a single report in the collection. It prefers the concept names that are frequent. IDF captures the frequency rate of a concept name across all the sets, i.e., how many sets in the collection contain a specific name. It prefers the concept names that are less frequent across the sets and are thus rare. Two variations of the measures—TF-IDF and Okapi BM25—were also selected because, theoretically, they can balance both types of preferences. The performances of these four measures were tested (see the "Method Verification" section).

The TPS and TSS are used to rank the candidate names based on how meaningful they are, because it is desirable for names that are meaningful to be ranked high. They are calculated based on the lexical patterns (i.e., lexical position and sequence) of terms in their original concept names. The TSSs are calculated based on the following lexical-position hypothesis: the contribution of a term's meaning to the entire meaning of a concept name decreases from right to left; the term on the rightmost side of a concept name contributes the most to the meaning of the name (Zhang and El-Gohary 2016). Thus, a candidate name that is mostly composed of terms from the right-hand side of an original name has a higher score than the name that is mostly composed of terms from the left-hand side. The TPS is calculated using Eq. (2), where *CICN* is a candidate identifier concept name, *OCN* is an original concept name in a set of original names *OCNs*, N is the number of names in the set, T is a term of *CICN*, M is the number of terms in *CICN*, $\mathrm{Index}_{\mathrm{OCN}}(T)$ is the index of T in *OCN*, and $|OCN|$ is the length of an original concept name:

$$\text{Term-position score}(CICN) = 1.0 + \frac{1}{N}\sum_{\{OCN\in OCNs\}}\frac{1}{M}\sum_{\{T\in CICN\}}\frac{\mathrm{Index}_{OCN}(T)}{|OCN|} \tag{2}$$

The TSSs are calculated based on the following lexical-sequence hypothesis: a candidate name with terms following the same consecutive sequence as they appear in its original name has a higher score. This hypothesis was made because using skip *n*-grams, although provides more candidate names with various term combinations, generates some names with terms that do not follow the same consecutive sequence and are, thus, generally less meaningful. For example, the terms of the skip bigram "asphalt wearing" do not follow the same consecutive sequence as they appear in the original concept name "asphalt deck wearing surface" and are, thus, less meaningful. The TSS is calculated using Eq. (3), where $I_{\{1,0\}} = 1.0$ if a candidate name has terms following the same consecutive sequence as they appear in an original name; otherwise, $I_{\{1,0\}} = 0$. The other notations follow those defined in Eq. (2):

$$\text{Term-sequence score}(CICN) = 1.0 + \frac{1}{N}\sum_{\{OCN\in OCNs\}} I_{\{1,0\}} \tag{3}$$

Identifier Concept Name Selection

Identifier concept name selection aims to select a final canonical identifier concept name from the top-ranking candidate names (one top-ranking name for each abstraction level). The selection is conducted hierarchically (i.e., in top-down fashion), so as to select an identifier with balanced abstraction and detailedness. For example, for a pair of top-ranking names in the adjacent abstraction levels (e.g., unigram and bigram names), if the detailed name fails to meet any of the *if* statements in a proposed selection rule, then the abstract name is selected as the final identifier; otherwise, the selection continues to the next pair (e.g., bigram and trigram names), until the abstract name in the pair is selected or no detailed name is available.

The proposed concept selection rule, which considers both the corpus statistics and the lexical patterns of the concept names, includes three cascading *if* statements:

- "If the ranking score of the detailed concept name added by an adjustment factor alpha is greater than the ranking score of the abstract concept name." The adjustment factor alpha is used to balance the abstraction and detailedness of the identifier concept names. A large value of alpha favors detailed names, and a small value favors abstract names.
- "If the word-association score of a detailed concept name is greater than a threshold value beta." The word-association score measures the degree to which two terms are related using corpus statistics (e.g., co-occurrence rates of the terms in a collection of inspection reports). The word-association score is used to make sure that the detailed concept names are lexical atoms (semantically coherent phrases, e.g., "map crack") rather than random combinations of terms. The normalized Google distance (Cilibrasi and Vitanyi 2007) was selected to calculate the word-association scores, because it is less negatively affected by extremely frequent terms, which cause lexical atoms to have low scores as random combinations. The threshold value beta is used to further balance the abstraction and detailedness. A large value of beta makes it more stringent for detailed names to be selected as identifiers and, thus, favors abstract names. A small value makes it easier for detailed names to be selected and, thus, favors such names.
- "If the part-of-speech (POS) pattern of the detailed concept name shows a noun-phrase pattern." The POS patterns (i.e., lexical class patterns of terms) are used to filter out candidate names that are not noun phrases, because a noun phrase is the most frequently occurring phrase type and is commonly used for naming concepts. Two noun-phrase patterns were selected and used: "noun + noun" and "adjective + noun," where "noun" could be a noun or a noun phrase.

As noted, two hyperparameters, alpha and beta, are used in the rule to balance the abstraction and detailedness of the identifier concept names. To find the optimized values for them, a total of 10,000 value combinations (with the values of alpha and beta ranging from 0 to 1 with a step size of 0.01, respectively) were tested. The combination with alpha = 0.38 and beta = 0.81 was empirically selected based on the testing results and was used for the experiments conducted in this research.

Entropy-Based Numerical Data Fusion

A new entropy-based numerical data fusion method is proposed. It fuses multiple numerical data that are complementary in describing the overall condition of a target into a single interval-based representation. In bridge inspection reports, multiple numerical data are the numerical measures of multiple instances, which are of the same type of deficiency but are at different locations on a bridge element. For example, in the report by WSDOT (2013), "5.5 meters" and "0.9 meters" (originally, "18 ft" and "3 ft" in the report) are the numerical measures of two "spall" instances that are at two different locations on the same "bridge rail." The proposed method includes four primary components: interval determination, degree tuple quantification, degree tuple fusion, and interval-based data representation. Fig. 3 shows the high-level algorithm for the proposed method.

Interval Determination

Interval determination aims to determine the number of intervals and the size of each interval for representing the fused data. Intervals are selected and used to represent the fused data in order to account for the uncertainty in the data and to avoid the exaggerated impact of minor fluctuations in continuous data on machine learning models. The proportional *k*-interval discretization method (Yang and Webb 2001) is used to define the intervals. This method was selected because it can use data to balance the tradeoff relationship between the number of intervals and the sizes of the intervals. A large number is preferred to capture more distinctive data instances for avoiding underfitting; at the same time, a large size is preferred to retain more data instances within an interval for avoiding overfitting. However, as the number increases, the size decreases.

Algorithm: Entropy-based numerical data fusion algorithm

1: **Input** All the unique numerical data instances in a dataset
// e.g., all the length measures of deck cracks in a set of bridge inspection reports
2: **Execute** Interval determination
// i.e., determine the number of intervals and the size of each interval using the proportional k-interval data discretization method
3: **Output** A set of intervals: $\{I_i\}, i = 1, \ldots, M$
// $M = K$ and the size of interval $I_i = K$, where K is the square root of the number of the unique numerical data instances (Line 1)
4: **Input** A set of numerical data instances to be fused
// e.g., all the length measures of deck cracks in a single report
5: **Execute** Degree tuple quantification, as per Eq. (4)
// i.e., quantify the membership, non-membership, and indeterminacy degree values of a data instance to interval I_i
6: **Output** Degree tuple matrix (*DTM*)
7: $DTM = \begin{bmatrix} DT_{1,1} & \ldots & DT_{1,N} \\ DT_{2,1} & \ldots & DT_{2,N} \\ \vdots & \ddots & \vdots \\ DT_{M,1} & \ldots & DT_{M,N} \end{bmatrix}$
// where $DT_{M,N}$ is a degree tuple containing the membership, non-membership, and indeterminacy degree values of data instance N to interval I_M, the degree values are calculating using Eq. (4), N is the number data instances, and M is the number of intervals
8: **Execute** Degree tuple fusion, as per Eq. (5)
// i.e., fuse the quantified degree tuples of an interval into a single tuple using Eq. (5)
9: **Output** Fused degree tuple vector (*FDTV*)
10: $FDTV = \begin{bmatrix} FDT_1 \\ FDT_2 \\ \vdots \\ FDT_M \end{bmatrix}$
// where FDT_M is a single degree tuple containing the fused membership, non-membership, and indeterminacy degree values of interval I_M, and the fused values are calculated using Eq. (5)
11: **Execute** Interval-based data representation
12: **For** each fused degree tuple $FDT_j, j = 1 \ldots M$
13: **Computed** the Euclidean distance between fused degree tuple and the ideal degree tuple [1, 0, 0]
// the tuple has a fused membership degree value 1 and the other two degree values 0, which means that the ideal interval corresponding to the ideal tuple can fully represent all the complementary data instances
14: **Output** The interval corresponding to the tuple with the smallest distance
15: **Output** Unified representation for the set of numerical data instances (Line 4): the count of data instances, the single representative interval of the data

Fig. 3. High-level algorithm for proposed entropy-based numerical data fusion method.

To balance this, the discretization method gives equal weight to them. The number of intervals is defined as $\sqrt{K}$, and the size of each interval is defined based on the minimum and maximum of the $\sqrt{K}$ unique data instances in the interval, where K is the number of unique data instances in a data set (e.g., unique deficiency measures of the same type of deficiency in a collection of inspection reports).

Degree Tuple Quantification

Degree tuple quantification aims to quantify the values contained in a degree tuple: membership, nonmembership, and indeterminacy degree values. Membership and nonmembership degrees are the extent of a data instance belonging and not belonging to an interval, respectively. Indeterminacy degree is the extent of hesitancy in claiming that the data instance belongs or does not belong to the interval. The normal cloud model (Li et al. 2009) is used to quantify these values because it can capture the uncertainty in the membership and nonmembership to allow for the modeling of the indeterminacy. The normal cloud model, which is based on the Gauss membership function and normal distribution, is a generalized normal distribution for quantifying the membership degree of a data instance belonging to an interval as a value between 0 and 1 (Li et al. 2009). The model assumes that the standard deviation of the Gauss membership function is not a fixed number but a random number following a normal distribution. Because of the randomness in drawing the standard deviation, for an interval, the Gauss function maps a data instance to many membership degree values (i.e., one-to-many mapping). Based on this mapping property, the authors propose to quantify the degree tuple using Eq. (4), where x is a data instance, I is an interval, $u_I(x)$ is a membership degree value mapped from the Gauss function, and *MDV*, *NDV*, and *IDV* are the membership, nonmembership, and indeterminacy degree values of x to I, respectively:

$$\begin{bmatrix} MDV_I(x) \\ NDV_I(x) \\ IDV_I(x) \end{bmatrix} = \begin{bmatrix} \min(\{u_I(x)\}) \\ 1 - \max(\{u_I(x)\}) \\ \max(\{u_I(x)\}) - \min(\{u_I(x)\}) \end{bmatrix} \quad (4)$$

Degree Tuple Fusion

Degree tuple fusion aims to fuse the quantified degree tuples of an interval into a single tuple. An information entropy–based fusion function is proposed to conduct the fusion. Information entropy is the average rate at which a stochastic process generates information (Shannon 1948); intuitively, it measures the amount of information in a random variable, where information entropy equal to zero indicates that the variable always generates the same information (Mehri and Darooneh 2011). Considering an interval as a variable that generates data instances, the information entropy of the interval is zero if it always generates the same particular instance (i.e., the membership degree value of this instance to the interval is 1) and cannot generate other instances (i.e., the membership degree values are 0). Conversely, the information entropy is greater than zero if the interval generates all instances (i.e., the membership degree values of these instances to the interval are between 0 and 1). In the first case, the interval can only represent the particular instance and is, thus, less representative of all the complementary data instances that collectively describe a target (e.g., the overall condition of a deficiency). As a result, its fused membership degree value should be downweighted and its nonmembership and indeterminacy degree values upweighted. In the second case, the interval can represent all the instances and is, thus, more representative. As a result, its fused membership degree value should be upweighted and the other two values downweighted. Based on the preceding analysis, the proposed information entropy–based function fuses the degree tuples of an interval as per Eq. (5), where W_I is the weight of the interval calculated using Eq. (6), i.e., the information entropy of the interval divided by the sum of information entropies of all the generated intervals. In Eqs. (5) and (6), $FMDV$, $FNDV$, and $FIDV$ are the fused membership, nonmembership, and indeterminacy degree values of interval I, respectively; N is the number of instances in the set of numerical data X; and M is the total number of intervals generated from the data discretization. The other notations follow those defined in Eq. (4):

$$\begin{bmatrix} FMDV_I(X) \\ FNDV_I(X) \\ FIDV_I(X) \end{bmatrix} = \begin{bmatrix} \dfrac{W_I}{N} \sum\limits_{\{x \in X\}} MDV_I(x) \\ \dfrac{1 - W_I}{N} \sum\limits_{\{x \in X\}} NDV_I(x) \\ \dfrac{1 - W_I}{N} \sum\limits_{\{x \in X\}} IDV_I(x) \end{bmatrix} \quad (5)$$

$$W_I = \frac{\sum_{i=1}^{N} MDV_I(x_i) \times \log_2 MDV_I(x_i)}{\sum_{j=1}^{M} \sum_{i=1}^{N} MDV_{I_j}(x_i) \times \log_2 MDV_{I_j}(x_i)} \quad (6)$$

Interval-Based Data Representation

Interval-based data representation aims to select an interval from all the possible intervals (defined in the "Interval Determination" section) for representing the multiple numerical data instances.

Table 1. Information on selected bridge inspection reports

Report No.	Region	Year	Number of records	Reference
1	Midwest	2006	409	MnDOT (2006)
2	Midwest	2015	152	MnDOT (2015)
3	Northeast	2008	97	CTDOT (2008)
4	Northeast	2013	451	CTDOT (2013)
5	Southeast	2008	163	LaDOTD (2008)
6	Southeast	2016	93	VDOT (2016)
7	Southwest	2007	88	NMDOT (2007)
8	Southwest	2008	32	NMDOT (2008)
9	West	2011	125	MDOT (2011)
10	West	2009	133	WSDOT (2009)

The selection is conducted based on the Euclidean distance between the fused degree tuple and the ideal degree tuple [1, 0, 0]. In the ideal degree tuple, the fused membership degree value is 1 and the other two degree values are 0. This means that the ideal interval corresponding to the ideal tuple can fully represent all the complementary data instances. Thus, an interval is selected if its fused degree tuple is the closest to the ideal tuple. As a result of the numerical data fusion, the multiple numerical data instances are represented in a unified way as follows: the count of the data instances, the single representative interval of the instances.

Method Verification

The verification aimed to evaluate the correctness of the proposed hybrid data fusion method. It included two steps: data set preparation and verification experiments. Two types of experiments were conducted to verify the two methods, respectively: named entity normalization experiments and numerical data fusion experiments.

Data Set Preparation

A data set, which includes 10 bridge inspection reports, was created. Information on these reports is summarized in Table 1. The information extraction methods of Liu and El-Gohary (2016, 2017b) were used to extract information about bridge conditions and maintenance actions from these reports and to represent the extracted information in a structured format. The data linking method of Liu and El-Gohary (2017a) was used to link the extracted data records that come from the same report and refer to the same entity. The linked records formed the data set for the normalization and fusion experiments.

Verification Experiments

Named Entity Normalization Experiments

The experiments aimed to implement the proposed normalization method to evaluate its accuracy by comparing the method-generated identifier concept names to the gold-standard identifiers. The method was implemented in a Python program (Python 2010). The natural language toolkit Porter stemmer and the "ngrams" function (Bird et al. 2009) were used for the morphological analysis and the n-gram generation, respectively. The Stanford POS tagger (Toutanova et al. 2003) was used to analyze the POS patterns of the concept names. The gold standard was prepared by human annotators—the first author and two other researchers with expertise in civil engineering, natural language processing, and machine learning. Full interannotator agreement was achieved after discussion. Accuracy, which is the ratio of the number of correct identifier concept names to the total number of identifier concept names, was calculated using Eq. (7):

$$\text{Accuracy} = \frac{\mathit{Number\ of\ correct\ identifier\ concept\ names}}{\mathit{Total\ number\ of\ identifier\ concept\ names}} \tag{7}$$

Numerical Data Fusion Experiments

The experiments aimed to implement the proposed fusion method to evaluate its stability in Monte Carlo simulations. Two factors could affect the stability of the method: the uncertainty in the data and the randomness in drawing the standard deviation of the Gauss function. Thus, two types of simulations were conducted: (1) simulations with data sampled from normal distributions, where each sampled data instance has an uncertainty level (i.e., the standard deviation of the normal distribution) ranging from 0.5 to 10 with a step size of 0.5, and (2) simulations with the times of randomly drawing the standard deviation ranging from 100 to 2,000 with a step size of 100. The number of iterations for each simulation run was set to 10,000. The method and simulations were implemented in a Python program (Python 2010). Information entropy was used to evaluate the stability of the method. It is equal to zero if, in a simulation run, the method can stably fuse the same set of multiple data instances and represent them using the same interval; otherwise, it increases. As a verification metric, it was calculated using Eq. (8), where M is the number of intervals, N_i is the times of the ith interval being selected to represent the same set of data instances, N is the number of iterations in a simulation run, and $N =$ 10,000 in this research:

$$\text{Information entropy} = -\sum_{i=1}^{M} \frac{N_i}{N} \times \log_2 \frac{N_i}{N} \tag{8}$$

Experimental Results and Discussion

Performance of Proposed Named Entity Normalization Method

Table 2 summarizes the performance results for the proposed normalization method. The results show that the method performed well: it achieved an average accuracy of 94.4%. Two important observations were also made based on the results.

First, the ranking function with the CSS, TPS, and TSS was effective. It achieved the highest accuracies of 85.4%, 89.3%, 100.0%, and 100.0% for bridge element, deficiency, maintenance action, and maintenance material names, respectively. But for deficiency cause names, the function with the CSS and TSS achieved the highest accuracy of 97.4%, which is 1.3% higher than that achieved using the function with all three. This is likely because the right-hand-side terms are not always the meaning-bearing terms in some deficiency cause names. For example, in the following names, the right-hand-side terms are less meaningful than those on the left-hand side: "debris buildup," "sand buildup," and "poor weld quality." The function without the TPS, thus, achieved a higher accuracy for deficiency cause names. Second, using POS patterns for selecting identifier concept names was effective. For example, using POS patterns achieved the highest accuracies of 85.4%, 97.4%, 100.0%, and 100.0% for the bridge element, deficiency cause, maintenance action, and maintenance material names, respectively. But for deficiency names, without using POS patterns achieved the highest accuracy of 89.3%, compared to 85.2% achieved using the "adjective + noun" and "noun + noun" patterns. This could be attributed to the fact that some deficiency names are not noun phrases (e.g., "pulled out" and "laterally misaligned"),

Table 2. Performance results for proposed named entity normalization method

Ranking function[a]	POS pattern	Accuracy for each concept name type				
		ET (%)	DY (%)	DC (%)	MA (%)	MM (%)
CSS	—	69.4	86.3	93.7	99.1	100.0
	Adj. + noun	64.6	54.9	61.7	99.1	79.3
	Noun + noun	73.3	80.5	89.9	99.1	100.0
	Adj. + noun and noun + noun	69.9	83.2	93.7	99.1	88.9
CSS × TPS	—	71.3	86.4	93.7	99.1	100.0
	Adj. + noun	64.7	54.9	61.7	99.1	79.3
	Noun + noun	75.1	80.6	89.9	99.1	100.0
	Adj. + noun and noun + noun	71.8	83.2	93.7	99.1	88.9
CSS × TSS	—	77.3	88.3	95.7	100.0	100.0
	Adj. + noun	74.3	55.7	63.8	100.0	79.3
	Noun + noun	82.6	81.1	93.5	100.0	100.0
	Adj. + noun and noun + noun	78.8	84.5	**97.4**	100.0	88.9
CSS × TPS × TSS	—	81.7	**89.3**	94.4	100.0	100.0
	Adj. + noun	75.2	55.8	63.8	100.0	79.3
	Noun + noun	**85.4**	81.8	92.3	**100.0**	**100.0**
	Adj. + noun and noun + noun	82.4	85.2	96.1	100.0	88.9

Note: POS = part-of-speech; Adj. = adjective; ET = bridge element; DY = deficiency; DC = deficiency cause; MA = maintenance action; and MM = maintenance material. The em dash (—) indicates that no POS pattern was used. The bold font indicates the highest accuracy for each concept name type.

[a]Four corpus-statistic measures for calculating CSS were tested: term frequency (TF), inverse document frequency (IDF), TF-IDF, and Okapi BM25. The performance results achieved using TF were reported because TF outperformed the others. The other measures use or partially use IDF, which frequently gave high scores to extremely rare concept names that should not be selected as identifiers.

and restricting identifiers to noun phrases led to a decrease in accuracy.

Performance of Proposed Numerical Data Fusion Method

Fig. 4 shows examples of the simulation results for fusing the deficiency length measures of patching on a girder: {304.8, 609.6, 609.6, 609.6, 609.6, 914.4, 1219.2}, where the unit is millimeter (originally, {12, 24, 24, 24, 24, 36, 48} in the report, where the unit is inch). The patterns of the simulation results for fusing the numerical data in the data set (as per Table 1) follow the same patterns as those shown in Fig. 4. Overall, the results show that the proposed fusion method was stable.

Two important observations were made based on the results. First, the fusion method was stable up to an uncertainty level of 2.0. As the uncertainty level increased from 2.0, the information entropy showed an increasing trend [Fig. 4(b)]. The increase in the information entropy indicates that the fusion method became unstable and started to represent the same set of deficiency measures using different intervals in a single simulation run [see the distributions of the intervals in Fig. 4(a)]. The uncertainties in the data negatively affect the quantification and the fusion of the degree values. Due to the uncertainties, these values changed in each fusion iteration of a simulation run, which made the fusion results of the same set of data vary. Second, the method was stable in the presence of the randomness of the standard deviation of the Gauss membership function. As shown in Fig. 4(d), increasing the randomness of the standard deviation (i.e., increasing the times of randomly drawing it) did not cause change in the information entropy. This indicates that the fusion method was stable and was able to represent the same set of numerical deficiency measures using the same interval in the simulations [see the distributions of the interval in Fig. 4(c)]. The standard deviation is bounded by a normal distribution in the cloud model (Li et al. 2009). Despite being random, the standard deviation is always within the bound, which made it not affect the stability of the fusion method.

Method Validation

The validation aimed to evaluate the performance of the proposed hybrid data fusion method in supporting its intended use—fusing data extracted from bridge inspection reports for supporting enhanced bridge deterioration prediction. It included two steps: data set preparation and validation experiments.

Data Set Preparation

The NBI data and the textual bridge inspection reports of 1,300 bridges, which are located in the state of Washington, were collected. The NBI data were collected from the Federal Highway Administration (FHWA 2019). The bridge inspection reports were collected from the Washington State Department of Transportation. Like the data set preparation for method verification, information extraction and data linking were conducted to process the reports. The linked records were then fused, as per Fig. 1, thereby forming a unified representation of the data extracted from the reports. Using the collected data, seven data sets were created. Table 3 summarizes the details of these data sets. In each data set, the data were split into a training data set and a testing data set. The training data set contains the 2013 data and the 2015 condition ratings of the decks, superstructures, and substructures of the bridges. The testing data set contains the 2015 data and the 2017 ratings. The 2015 ratings were used as the target classes for training the prediction models, and the 2017 ratings were used as the gold standard for testing the performance of the models.

Validation Experiments

The validation experiments aimed to develop machine learning models for predicting the future condition ratings of decks, superstructures, and substructures. The decision tree algorithm was selected from among other existing learning algorithms, such as Naïve Bayes (Maron 1961), support vector machines (Cortes and Vapnik 1995),

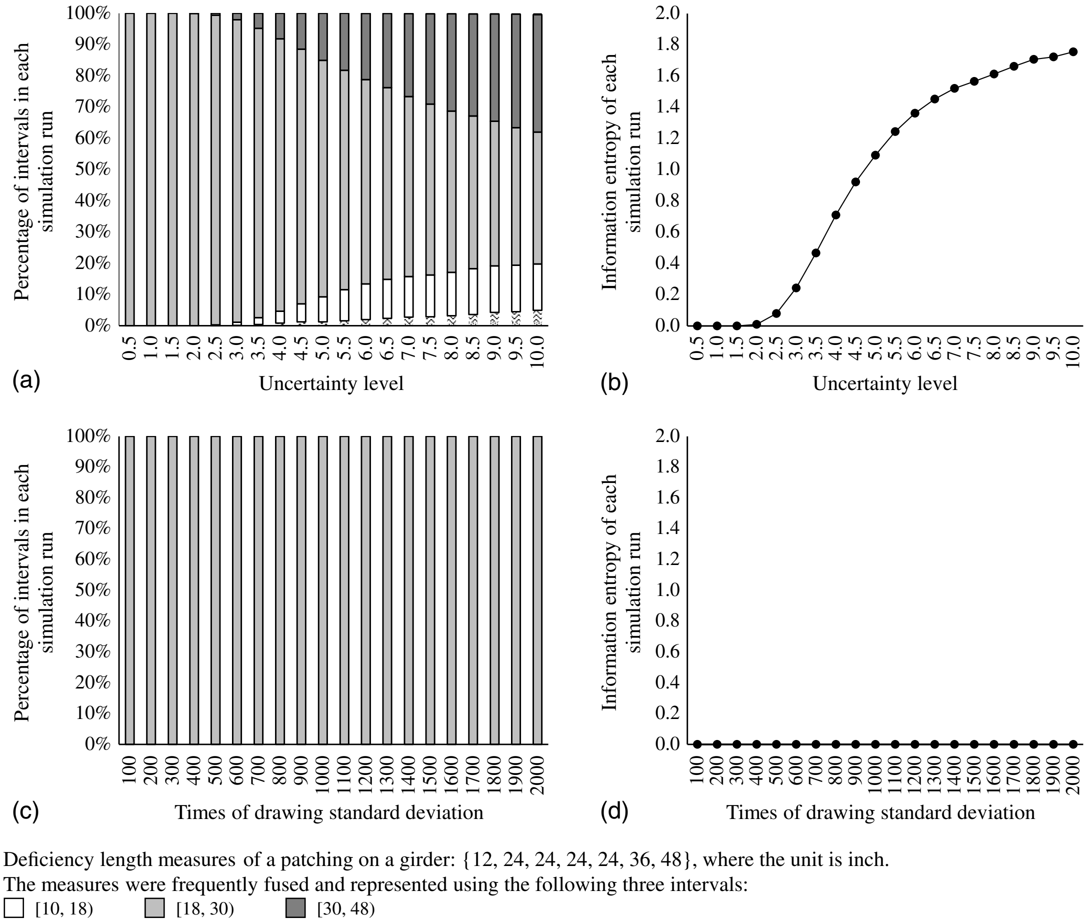


Fig. 4. Examples of Monte Carlo simulation results for fusing multiple deficiency measures: (a) percentage of intervals in each simulation run against uncertainty level; (b) information entropy of each simulation run against uncertainty level; (c) percentage of intervals in each simulation run against times of drawing standard deviation; and (d) information entropy of each simulation run against times of drawing standard deviation.

and neural networks and deep learning (e.g., LeCun et al. 2015), to develop the models. This is because a tree-based algorithm can directly handle both categorical and numerical features without the need for one-hot encoding (Gupta et al. 2017). One-hot encoding transforms categorical features into numerical features using dummy variables. Such variables increase the dimensionality and the sparsity of the feature space, which negatively affects the performance of machine learning models (Guo and Berkhahn 2016) and would negatively affect validation. Seven main types of prediction models were developed, with each type trained and tested using the data in one of the data sets, as per Table 3. Average accuracy was selected as the validation metric because it can capture both systematic error and random error (ISO 1994), which is important to validate if learning from the fused data can reduce both types of errors in order to improve the accuracy. Average accuracy is the average of the ratio of the number of correctly predicted condition ratings to the total number of ratings per condition rating category. It was calculated using Eq. (9), where N is the number of condition rating categories, *CRs* are condition ratings, and *CRC* is a condition rating category; and, for the ith condition rating category, TP (true positive) is the number of condition ratings correctly predicted as this category, FP (false positive) is the number condition ratings incorrectly predicted as this category, and FN (false negative) is the number of condition ratings that should have been but were not predicted as this category:

$$\text{Average accuracy} = \frac{1}{N}\sum_{i=1}^{N}\frac{TP_i}{TP_i + FP_i + FN_i} = \frac{1}{N}\sum_{i=1}^{N}\frac{Number\ of\ CRs\ correctly\ predicted\ as\ ith\ CRC}{Total\ number\ of\ CRs\ in\ ith\ CRC} \tag{9}$$

Experimental Results and Discussion

The performance results for predicting the future condition ratings of decks, superstructures, and substructures are presented in Fig. 5. Overall, the results show that the proposed hybrid data fusion method was effective in fusing the data extracted from bridge inspection reports for supporting enhanced bridge deterioration prediction. Three important observations were also made based on the results.

First, learning from textual bridge inspection reports, in addition to NBI data, was able to improve the prediction performance. Learning from both NBI data and the report data fused by the proposed method, compared to learning from NBI data alone, improved the prediction accuracies for decks, superstructures, and substructures by 8.3%, 9.5%, and 8.3%, respectively. NBI data, which mainly describe condition ratings and as-built characteristics

Table 3. Summary of created data sets

Data set	Data[a]	Purpose[b]
1	NBI data	Used to develop baseline prediction models to evaluate whether further learning from the data extracted from bridge inspection reports would be able to improve the performance of bridge deterioration prediction
2	NBI data + unfused report data	Used to develop baseline prediction models to evaluate whether the fusion of the data extracted from the reports is needed in order to improve the performance of bridge deterioration prediction
3	NBI data + fused report data (fused by the proposed hybrid data fusion method)	Used to develop prediction models to evaluate the performance of the proposed hybrid data fusion method in fusing the data extracted from the reports for supporting enhanced bridge deterioration prediction
4	NBI data + fused report data (where deficiency measures were fused by taking the maximum of the measures, i.e., using the worst deterioration case)	Used to develop baseline prediction models to evaluate whether the deficiency measures fused by the proposed method could better support the prediction compared to the measures fused by taking the maximum
5	NBI data + fused report data (where deficiency measures were fused using one of the central tendency measures, including arithmetic mean, Bonferroni mean, geometric mean, harmonic mean, Heroin mean, power mean, median, and mode)	Used to develop baseline prediction models to evaluate whether the deficiency measures fused by the proposed method could better support the prediction compared to the measures fused using the mean (or the total, i.e., the multiple measures were represented by two values: number of measure instances and mean of measures)
6	NBI data + fused report data (where deficiency measures were fused using one of the variation measures, including range, mean absolute difference, coefficient of variation, standard deviation, and variance)	Used to develop baseline prediction models to evaluate whether the deficiency measures fused by the proposed method could better support the prediction compared to the measures fused using the variation
7	NBI data + fused report data (where same set of deficiency measures were fused using one of the central tendency measures and using one of the variation measures, i.e., the same set was represented twice)	Used to develop baseline prediction models to evaluate whether the deficiency measures fused by the proposed method could better support the prediction compared to the measures fused using the combinations of the central tendency and the variation measures (e.g., the measures were represented using three values: number of measure instances, arithmetic mean of measures, and variance of measures)

Note: NBI = National Bridge Inventory.

[a]The concept names in the textual bridge inspection reports in Data sets 4–7 were fused by the proposed unsupervised named entity normalization method.

[b]Bridge deterioration prediction means the predictions of future condition ratings of decks, superstructures, and substructures.

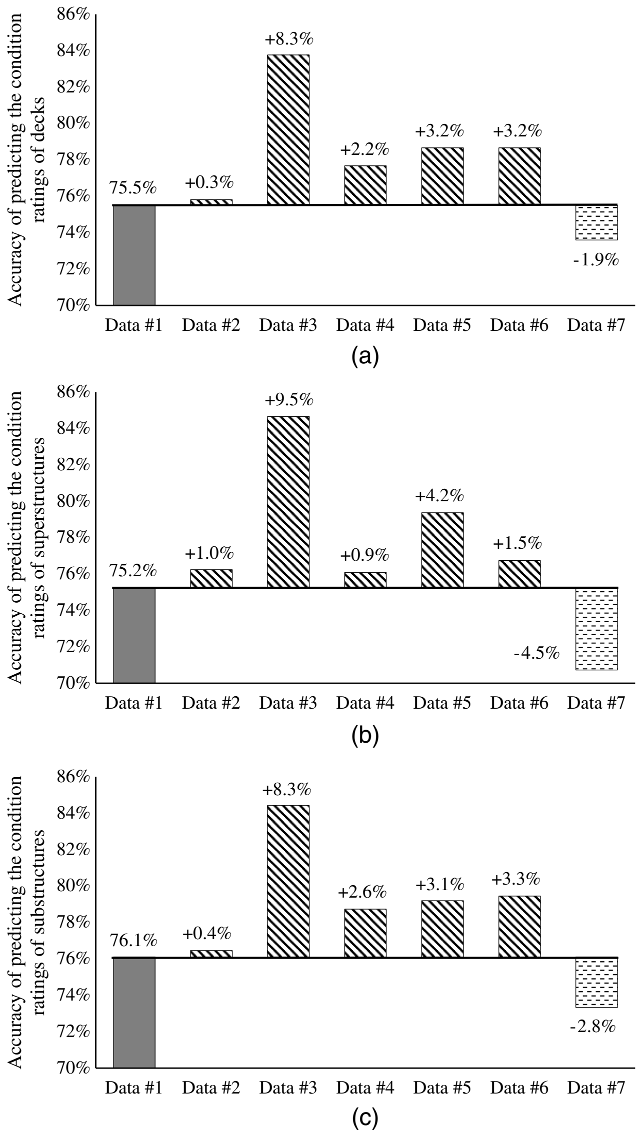
of bridges, are certainly important, but they do not include descriptions about the element-level deterioration conditions of bridges, such as those in bridge inspection reports. Such descriptions are much more detailed and dynamic in capturing the deterioration conditions of bridges in each inspection year and are, therefore, more informative in capturing the patterns of how the condition ratings evolve over time. Hence, they helped improve the performance of predicting future ratings.

Second, data fusion is very important for learning from the data extracted from bridge inspection reports to improve prediction performance. Learning from the unfused report data was only able to marginally improve the prediction accuracies by 0.3%, 1.0%, and 0.4%, respectively. But learning from the report data fused by the proposed method improved these accuracies, highly, by 8.7% on average. The multiple concept names and numerical data in inspection reports negatively affected the generalizability of the machine learning models, which limited the performance of learning from bridge inspection reports in improving the prediction performance.

Third, the proposed entropy-based data fusion method was effective in fusing the numerical data in the reports for improving the prediction performance. Learning from the data fused by taking the maximum (i.e., using the measure corresponding to the worse deterioration case), the mean/total, or the variation was only able to improve the prediction accuracy by 2.7% on average. Learning from the data fused by the combinations of the means and the variations, even, decreased the accuracy by around 3.0%, due to the doubled size of the feature dimensionality caused by the combinations. The proposed method improved the accuracy by 8.7% on average, which is significantly higher than the improvement rates achieved using the other methods. This is largely attributed to the fact that the proposed method uses data discretization to define the interval-based representation of the fused data and utilizes information entropy to fuse data into a single representative interval. Thus, the proposed method takes balancing the overfitting and underfitting of the machine learning prediction model and the complementarity of the data into account, resulting in an improved prediction performance.

Limitations

Three main limitations of this research are acknowledged. First, the unsupervised named entity normalization method uses the normalized Google distance to assess the associations of words for selecting identifier concept names. Half of the normalization errors were caused by the incorrectly assessed associations, where the total error rate of the method is 5.6%. This distance was developed mainly to assess the associations of words in general-domain texts. In future research, the authors plan to develop a word-association measure that can better adapt to domain-specific text and test its impact on named entity normalization. Second, the entropy-based data fusion method focuses on fusing data from a single type of source (e.g., deficiency measures from the text). It requires modifications and/or extensions when used for fusing multimodal data (e.g., deficiency measures from text, images, and sensors). Also in future research, the authors plan to develop a multilevel context-based fusion method that can capture the context of data (e.g., the types of sensing devices and their reliabilities) for supporting multimodal data fusion. Third, the bridge deterioration prediction models were developed using the decision tree algorithm. Although this algorithm is suitable for the purpose of validating the hybrid data fusion method, it is limited in dealing with highly dimensional and imbalanced data such as bridge data and is thus limited in learning from



Note: The types of the data (data #1 to #7) are defined in Table 3. The accuracies for data #5 to #7 are the averages across different measures or combinations (see Table 3).

Fig. 5. Performance results for predicting future condition ratings of (a) decks; (b) superstructures; and (c) substructures.

the data to better predict the deterioration. Other learning algorithms such as support vector machines and neural networks, although are effective in capturing the non-linearity and complex patterns in data, also suffer from these limitations when used as-is (Gao et al. 2017; Kaur et al. 2019). Thus, in their future work, the authors plan to study how to integrate dimensionality reduction, data sampling, and machine learning algorithms to address these limitations in order to more effectively learn from the data for enhanced data-driven bridge deterioration prediction.

Contributions to Body of Knowledge

This research contributes to the body of knowledge in three primary ways. First, it offers a new unsupervised named entity normalization method for fusing concept names without human involvement. The method captures both surface-form and abstraction-detailedness variations in concept names to fuse them into canonical identifier concept names with balanced abstraction and detailedness. It thus extends the state of the art in named entity normalization, where most of the existing methods rely heavily on human-developed dictionaries/data and can only capture surface-form variations to fuse concept names into their canonical forms. Second, this research offers a new entropy-based data fusion method. The method uses data discretization to define the interval-based representation of the fused data and leverages information entropy to fuse data that are complementary into a single representation. It thus adds to the state of the art in numerical data fusion, where most of the existing methods focus on fusing data that are conflicting or imprecise. Third, this research allows for better learning from the data extracted from textual bridge inspecting reports for enhanced deterioration prediction. Inspection report data are highly complex because they include multiple concept names and numerical measures to describe the same entity. Directly learning from the extracted data, without fusion, is limited in improving the performance of deterioration prediction, and learning from improperly fused data could even compromise the performance. This research—by proposing a hybrid data fusion method—allows for fusing data extracted from textual reports for enhanced deterioration prediction performance. It thus offers new knowledge on how to effectively use data that have varying levels of detailedness and have complementary characteristics in data analytics for supporting data-driven applications. The gained knowledge would be critical to fusing and learning from multisource heterogeneous bridge data—not only textual data from inspection reports but also structured bridge inventory data and unstructured bridge condition data from images and sensors.

Conclusions and Future Work

In this paper, the authors proposed a new hybrid data fusion method for fusing data extracted from textual bridge inspection reports for supporting enhanced bridge deterioration prediction. At the cornerstone of the proposed method are two submethods, an unsupervised named entity normalization method and an entropy-based numerical data fusion method, for fusing concept names and numerical data, respectively. A set of experiments were conducted to evaluate the performance of the proposed method. Two important conclusions were drawn from the experimental results. First, learning from the report data, in addition to NBI data, was able to improve the prediction performance. Compared to learning from NBI data alone, further learning from the report data fused by the proposed method improved the accuracies for predicting the future condition ratings of decks, superstructures, and substructures by 8.3%, 9.5%, and 8.3%, respectively. Second, the proposed data fusion method was effective in fusing data extracted from the reports for supporting enhanced deterioration prediction. Compared to learning from the unfused report data, learning from the report data fused by the proposed method improved the prediction accuracies by 8.0%, 8.5%, and 7.9%, respectively. The experimental results show the promise of the proposed method, where researchers and practitioners in the civil infrastructure domain could use heterogeneous data from multiple sources (especially textual data) in a fused way, rather than using data of a single type or from a single source in isolation, for enhanced data-driven predictive analytics and decision-making.

In future research, the authors plan to focus their research efforts on two main directions. First, the authors will develop new machine learning algorithms that are able to effectively deal with highly dimensional and imbalanced data such as bridge data. Different embedding methods for reducing data dimensionality and sampling methods for balancing data distributions will be studied.

Their performances in supporting the analysis of fused textual bridge inspection data will be tested. Using the developed algorithms, further learning from fused data—compared to only learning from NBI data—is expected to show an even more significant improvement in the deterioration prediction performance. Second, the authors will use the developed algorithms to learn from multisource heterogeneous bridge data, in order to allow for the predictions of the types and quantities of deficiencies that a bridge element could develop in the future, not only the condition ratings of bridges. Such bridge data would include NBI and National Bridge Element (NBE) data, textual bridge inspection data, data about bridge deficiencies and conditions from images and sensors, traffic data, and weather data. These efforts would create new knowledge on how to use heterogeneous bridge data from multiple sources in an analyzed and integrative manner to better understand and predict bridge deterioration for enhanced data-driven maintenance decision-making.

Data Availability Statement

Some or all data, models, or code generated or used during the study are available in a repository or online in accordance with funder data retention policies [the bridge inspection reports as per Table 1 and the National Bridge Inventory data from the Federal Highway Administration (FHWA 2019)]. Some or all data, models, or code used during the study were provided by a third party (the bridge inspection reports from the Washington State Department of Transportation). Direct requests for these materials may be made to the provider as indicated in the acknowledgments. Some or all data, models, or code generated or used during the study are available from the corresponding author by request (the Python code developed for the implementation and the experimental testing of the proposed hybrid data fusion method).

Acknowledgments

The authors would like to thank the National Science Foundation (NSF). This paper is based upon work supported by NSF under Grant No. 1937115. Funding for this research was also provided in part by the National Center for Supercomputing Applications (NCSA) at the University of Illinois at Urbana-Champaign through the NCSA Faculty Fellows program. The authors would also like to thank the Washington State Department of Transportation for providing access to the bridge inspection reports. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the aforementioned entities.

References

AASHTO. 2010. *AASHTO bridge element inspection guide manual*. Washington, DC: AASHTO.

AASHTO. 2018. "AASHTOWare Bridge Management." Washington, DC. Accessed March 29, 2020. https://www.aashtoware.org/wp-content/uploads/2018/03/Bridge-Management-Product-Brochure-FY-2019-01192019.pdf.

Akhtar, M. S., U. K. Sikdar, and A. Ekbal. 2015. "IITP: Hybrid approach for text normalization in Twitter." In *Proc., ACL 2015 Workshop on Noisy User-generated Text*, 106–110. Stroudsburg, PA: Association for Computational Linguistics.

Alhelbawy, A., and R. Gaizauskas. 2014. "Graph ranking for collective named entity disambiguation." In *Proc., 52nd Annual Meeting of the Association for Computational Linguistics*, 75–80. Stroudsburg, PA: Association for Computational Linguistics.

Bird, S., E. Klein, and E. Loper. 2009. *Natural language processing with Python: Analyzing text with the natural language toolkit.* Beijing: O'Reilly Media.

Boström, H., S. F. Andler, M. Brohede, R. Johansson, A. Karlsson, J. Van Laere, L. Niklasson, M. Nilsson, A. Persson, and T. Ziemke. 2007. *On the definition of information fusion as a field of research.* Technical Rep. No. HS-IKI-TR-07-006. Skövde, Sweden: Univ. of Skövde.

Chang, M., M. Maguire, and Y. Sun. 2018. "Stochastic modeling of bridge deterioration using classification tree and logistic regression." *J. Infrastruct. Syst.* 25 (1): 04018041. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000466.

Church, K., and P. Hanks. 1990. "Word association norms, mutual information, and lexicography." *Comput. Ling.* 16 (1): 22–29. https://doi.org/10.3115/981623.981633.

Cilibrasi, R. L., and P. M. Vitanyi. 2007. "The Google similarity distance." *IEEE Trans. Knowl. Data Eng.* 19 (3): 370–383. https://doi.org/10.1109/TKDE.2007.48.

Cortes, C., and V. Vapnik. 1995. "Support-vector networks." *Mach. Learn.* 20 (3): 273–297. https://doi.org/10.1007/BF00994018.

CTDOT (Connecticut Department of Transportation). 2008. "Bridge safety inspection: Sunnyside Street over Yantic River." Newington, CT: CTDOT. Accessed December 18, 2019. https://portal.ct.gov/-/media/DOT/documents/dbridgepubs/BSE04746IR20080923INDEPTHpdfpdf.pdf?la=en.

CTDOT (Connecticut Department of Transportation). 2013. "Bridge inspection report BRI-18." Newington, CT: CTDOT. Accessed December 18, 2019. https://portal.ct.gov/-/media/DOT/documents/dbridgepubs/.

Destercke, S., D. Dubois, and E. Chojnacki. 2008. "Possibilistic information fusion using maximal coherent subsets." *IEEE Trans. Fuzzy Syst.* 17 (1): 79–92. https://doi.org/10.1109/TFUZZ.2008.2005731.

Fakhraei, S., and J. L. Ambite. 2018. "NSEEN: Neural semantic embedding for entity normalization." Preprint, submitted November 19, 2015. http://arxiv.org/abs/1811.07514.

FHWA (Federal Highway Administration). 1995. *Recoding and coding guide for the structure inventory and appraisal of the nation's bridges.* FHWA-PD-96-01. Washington, DC: FHWA.

FHWA (Federal Highway Administration). 2013. "Summary report on the FHWA LTBP workshop to identify bridge substructure." Washington, DC: FHWA. Accessed March 29, 2020. https://www.fhwa.dot.gov/publications/research/infrastructure/structures/ltbp/11037/11037.pdf.

FHWA (Federal Highway Administration). 2014. "FHWA LTBP bridge performance primer." Washington, DC: FHWA. Accessed March 29, 2020. https://www.fhwa.dot.gov/publications/research/infrastructure/structures/ltbp/14032/14032.pdf.

FHWA (Federal Highway Administration). 2019. "National bridge inventory." Washington, DC: FHWA. Accessed June 21, 2019. https://www.fhwa.dot.gov/bridge/nbi/ascii.cfm.

FHWA (Federal Highway Administration). 2020. "LTBP deterioration condition forecast models." Washington, DC: FHWA. Accessed March 29, 2020. https://cms7.fhwa.dot.gov/research/long-term-infrastructure-performance/ltbp/data-analysis.

Gao, L., J. Song, X. Liu, J. Shao, J. Liu, and J. Shao. 2017. "Learning in high-dimensional multimedia data: The state of the art." *Multimedia Syst.* 23 (3): 303–313. https://doi.org/10.1007/s00530-015-0494-1.

Guo, C., and F. Berkhahn. 2016. "Entity embeddings of categorical variables." Preprint, submitted April 24, 2015. http://arxiv.org/abs/1604.06737.

Gupta, B., A. Rawat, A. Jain, A. Arora, and N. Dhami. 2017. "Analysis of various decision tree algorithms for classification in data mining." *Int. J. Comput. Appl.* 163 (8): 15–19. https://doi.org/10.5120/ijca2017913660.

Han, Y., W. Chen, X. Xiong, Q. Li, Z. Qiu, and T. Wang. 2019. "Wide & deep learning for improving named entity recognition via text-aware named entity normalization." In *Proc., AAAI 2019 Workshop on Recommender Systems and Natural Language Processing*. Menlo Park, CA: Association for the Advancement of Artificial Intelligence.

He, Z., W. Jiang, and F. T. Chan. 2018. "Evidential supplier selection based on interval data fusion." *Int. J. Fuzzy Syst.* 20 (4): 1159–1171. https://doi.org/10.1007/s40815-017-0426-7.

Huang, Y. H. 2010. "Artificial neural network model of bridge deterioration." *J. Perform. Constr. Facil.* 24 (6): 597–602. https://doi.org/10.1061/(ASCE)CF.1943-5509.0000124.

ISO. 1994. *Accuracy (trueness and precision) of measurement methods and results. Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method.* ISO 5725-2. Geneva: ISO.

Jiang, J. J., and D. W. Conrath. 1997. "Semantic similarity based on corpus statistics and lexical taxonomy." Preprint, submitted September 20, 1997. http://arxiv.org/abs/cmp-lg/9709008.

Jiang, W., C. Xie, M. Zhuang, Y. Shou, and Y. Tang. 2016. "Sensor data fusion with Z-numbers and its application in fault diagnosis." *Sensors* 16 (9): 1509. https://doi.org/10.3390/s16091509.

Jin, N. 2015. "NCSU-SAS-Ning: Candidate generation and feature engineering for supervised lexical normalization." In *Proc., ACL 2015 Workshop on Noisy User-Generated Text*, 87–92. Stroudsburg, PA: Association for Computational Linguistics.

Kaur, H., H. S. Pannu, and A. K. Malhi. 2019. "A systematic review on imbalanced data challenges in machine learning: Applications and solutions." *ACM Comput. Surv.* 52 (4): 1–36. https://doi.org/10.1145/3343440.

Khaleghi, B., A. Khamis, F. O. Karray, and S. N. Razavi. 2013. "Multisensor data fusion: A review of the state-of-the-art." *Inf. Fusion* 14 (1): 28–44. https://doi.org/10.1016/j.inffus.2011.08.001.

LaDOTD (Louisiana Department of Transportation and Development). 2008. "Bridge inspection report: Inspection of the Hale Boggs Memoril Bridge in Luling-Destrehan." Baton Rouge, LA: LaDOTD. Accessed December 18, 2019. http://www8.dotd.la.gov/Proposals/450-37-0022/Reports%20and%20Notes/Bridge%20Inspection%20Report.pdf.

Lahat, D., T. Adali, and C. Jutten. 2015. "Multimodal data fusion: An overview of methods, challenges, and prospects." *Proc. IEEE* 103 (9): 1449–1477. https://doi.org/10.1109/JPROC.2015.2460697.

Le, T., and H. D. Jeong. 2017. "NLP-based approach to semantic classification of heterogeneous transportation asset data terminology." *J. Comput. Civ. Eng.* 31 (6): 04017057. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000701.

LeCun, Y., Y. Bengio, and G. Hinton. 2015. "Deep learning." *Nature* 521 (7553): 436–444. https://doi.org/10.1038/nature14539.

Lee, H. C., Y. Y. Hsu, and H. Y. Kao. 2016. *AuDis: An automatic CRF-enhanced disease normalization in biomedical text*. Oxford, UK: Oxford University Press.

Leeman-Munk, S., J. Lester, and J. Cox. 2015. "NCSU_SAS_SAM: Deep encoding and reconstruction for normalization of noisy text." In *Proc., ACL 2015 Workshop on Noisy User-generated Text*, 154–161. Stroudsburg, PA: Association for Computational Linguistics.

Li, D., C. Liu, and W. Gan. 2009. "A new cognitive model: Cloud model." *Int. J. Intell. Syst.* 24 (3): 357–375. https://doi.org/10.1002/int.20340.

Li, T., and D. Harris. 2019. "Automated construction of bridge condition inventory using natural language processing and historical inspection reports." In *Proc., Nondestructive Characterization and Monitoring of Advanced Materials, Aerospace, Civil Infrastructure, and Transportation XIII*, 109710T1–109710T8. Bellingham, WA: International Society for Optics and Photonics.

Liu, K., and N. El-Gohary. 2016. "Ontology-based sequence labelling for automated information extraction for supporting bridge data analytics." *Procedia Eng.* 145: 504–510. https://doi.org/10.1016/j.proeng.2016.04.035.

Liu, K., and N. El-Gohary. 2017a. "Ontology-based data integration for supporting big bridge data analytics." In *Proc., 2017 CSCE Annual Conf.* Montreal: Canadian Society for Civil Engineering.

Liu, K., and N. El-Gohary. 2017b. "Ontology-based semi-supervised conditional random fields for automated information extraction from bridge inspection reports." *Autom. Constr.* 81 (Sep): 313–327. https://doi.org/10.1016/j.autcon.2017.02.003.

Liu, K., and N. El-Gohary. 2018. "Unsupervised named entity normalization for supporting information fusion for big bridge data analytics." In *Proc., Workshop of the European Group for Intelligent Computing in Engineering*. Cham, Switzerland: Springer.

Liu, K., and N. El-Gohary. 2019. "A hybrid information fusion method for fusing data extracted from inspection reports for supporting bridge data analytics." In *Proc., ASCE International Conference on Computing in Civil Engineering*. Reston, VA: ASCE.

Liu, X., M. Zhou, F. Wei, Z. Fu, and X. Zhou. 2012. "Joint inference of named entity recognition and normalization for tweets." In *Proc., 50th Annual Meeting of the Association for Computational Linguistics*, 526–535. Stroudsburg, PA: Association for Computational Linguistics.

Lu, P., H. Wang, and D. Tolliver. 2019. "Prediction of bridge component ratings using ordinal logistic regression model." In *Mathematical problems in engineering*. London: Hindawi.

Magdy, W., K. Darwish, O. Emam, and H. Hassan. 2007. "Arabic cross-document person name normalization." In *Proc., 2007 Workshop on Computational Approaches to Semitic Languages: Common Issues and Resources*, 25–32. Stroudsburg, PA: Association for Computational Linguistics.

Mann, P. S. 1995. "Introductory statistics." *J. R. Stat. Soc.* 158 (2): 339.

Maron, M. E. 1961. "Automatic indexing: An experimental inquiry." *J. ACM* 8 (3): 404–417. https://doi.org/10.1145/321075.321084.

MDOT (Montana Department of Transportation). 2011. "Bridge inspection report: Heron Truss Bridge over Cabinet Gorge Reservoir." Helena, MT: MDOT. Accessed December 18, 2019. http://ftp://ftp.mdt.mt.gov/contract/contractors/CABINET_GORGE_1_M_W_HERON/INSP_FINAL_HERON_REPORT_2011.pdf.

Mehri, A., and A. H. Darooneh. 2011. "The role of entropy in word ranking." *Physica A* 390 (18–19): 3157–3163. https://doi.org/10.1016/j.physa.2011.04.013.

Mendenhall, W. M., and T. L. Sincich. 2016. *Statistics for engineering and the sciences*. Boca Raton, FL: Chapman & Hall/CRC Press.

Mihalcea, R., and A. Csomai. 2007. "Wikify!: Linking documents to encyclopedic knowledge." In *Proc., 16th Conf. on Information and Knowledge Management*, 233–242. New York: Association for Computing Machinery.

MnDOT (Minnesota Department of Transportation). 2006. "Fracture critical bridge inspection in-depth report: I-35W over the Mississippi River at Minneapolis, MN." Saint Paul, MN: MnDOT. Accessed December 18, 2019. https://www.leg.state.mn.us/docs/2007/other/070549.pdf.

MnDOT (Minnesota Department of Transportation). 2015. "2015 routine and fracture critical bridge inspection report: TH 194 NB over Superior St." Saint Paul, MN: MnDOT. Accessed December 19, 2019. http://www.dot.state.mn.us/consult/documents/notices/1026462-BR%2069839%20FC%20REPORT%20-%202015.pdf.

Mohammadi, A., S. H. Javadi, D. Ciuonzo, V. Persico, and A. Pescapè. 2019. "Distributed detection with fuzzy censoring sensors in the presence of noise uncertainty." *Neurocomputing* 351 (Jul): 196–204. https://doi.org/10.1016/j.neucom.2019.03.044.

NMDOT (New Mexico Department of Transportation). 2007. "Bridge inspection report: 000000000006487." Santa Fe, NM: NMDOT. Accessed December 18, 2019. http://santafempo.org/wp-content/uploads/2009/07/Bridge-Inspection-Reports.pdf.

NMDOT (New Mexico Department of Transportation). 2008. "Bridge inspection report: 000000000006549." Santa Fe, NM: NMDOT. Accessed December 18, 2019. http://santafempo.org/wp-content/uploads/2009/07/Bridge-Inspection-Reports.pdf.

NYSDOT (New York State Department of Transportation). 2015. "BIN: 1018259 Bridge inspection report." Albany, NY: NYSDOT. Accessed March 29, 2020. https://www.dot.ny.gov/main/business-center/designbuildproject20/repository/2015%20Bridge%20Inspection%20Reporte%20-%2020151228.pdf.

Pafilis, E., S. P. Frankild, L. Fanini, S. Faulwetter, C. Pavloudi, A. Vasileiadou, C. Arvanitidis, and L. J. Jensen. 2013. "The SPECIES and ORGANISMS resources for fast and accurate identification of taxonomic names in text." *PLOS One* 8 (6): e65390. https://doi.org/10.1371/journal.pone.0065390.

Pawlak, Z. 2012. *Rough sets: Theoretical aspects of reasoning about data.* Berlin: Springer.

Python. 2010. *Python 2.7 documentation*. Wilmington, DE: Python Software Foundation.

Sentz, K., and S. Ferson. 2002. *Combination of evidence in Dempster-Shafer theory.* Sandia Rep. No. SAND2002-0835. Albuquerque, NM: Sandia National Laboratories.

Shafer, G. 1976. *A mathematical theory of evidence.* Princeton, NJ: Princeton University Press.

Shannon, C. E. 1948. “A mathematical theory of communication.” *Bell Syst. Tech. J.* 27 (3): 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x.

Song, C., H. Zhao, Z. Xu, and Z. Hao. 2019. “Interval-valued probabilistic hesitant fuzzy set and its application in the Arctic geopolitical risk evaluation.” *Int. J. Intell. Syst.* 34 (4): 627–651. https://doi.org/10.1002/int.22069.

Tahmasebi, A. M., H. Zhu, G. Mankovich, P. Prinsen, P. Klassen, S. Pilato, R. van Ommering, P. Patel, M. L. Gunn, and P. Chang. 2019. “Automatic normalization of anatomical phrases in radiology reports using unsupervised learning.” *J. Digital Imaging* 32 (1): 6–18. https://doi.org/10.1007/s10278-018-0116-5.

Tian, W., P. de Wilde, Z. Li, J. Song, and B. Yin. 2018. “Uncertainty and sensitivity analysis of energy assessment for office buildings based on Dempster-Shafer theory.” *Energy Convers. Manage.* 174 (Oct): 705–718. https://doi.org/10.1016/j.enconman.2018.08.086.

Torra, V. 2010. “Hesitant fuzzy sets.” *Int. J. Intell. Syst.* 25 (6): 529–539. https://doi.org/10.1002/int.20418.

Toutanova, K., D. Klein, C. D. Manning, and Y. Singer. 2003. “Feature-rich part-of-speech tagging with a cyclic dependency network.” In *Proc., 2003 Conf. of the North American Chapter of the Association for Computational Linguistics on Human Language Technology*, 173–180. Stroudsburg, PA: Association for Computational Linguistics.

Varga, O., I. G. Nagy, P. Burai, T. Tomor, C. Lénárt, and S. Szabó. 2018. “Land cover analysis based on descriptive statistics of Sentinel-2 time series data.” *Landscape Environ.* 12 (2): 1–9. https://doi.org/10.21120/LE/12/2/1.

VDOT (Virginia Department of Transportation). 2016. “Bridge inspection report: Forest Mills Road over tributary of crooked run.” Richmond, VA: VDOT. Accessed December 18, 2019. https://healthdocbox.com/83825408-Dental_Care/Virginia-department-of-transportation-northern-virginia-district.html.

Wagner, J., and J. Foster. 2015. “DCU-ADAPT: Learning edit operations for microblog normalisation with the generalised perceptron.” In *Proc., ACL 2015 Workshop on Noisy User-Generated Text*, 93–98. Stroudsburg, PA: Association for Computational Linguistics.

Wang, F., W. Wu, Z. Li, and M. Zhou. 2017. “Named entity disambiguation for questions in community question answering.” *Knowl.-Based Syst.* 126 (Jun): 68–77. https://doi.org/10.1016/j.knosys.2017.03.017.

Wei, Q., T. Chen, R. Xu, Y. He, and L. Gui. 2016. *Disease named entity recognition by combining conditional random fields and bidirectional recurrent neural networks*. Oxford, UK: Oxford University Press.

Wimmer, M., B. Schuller, D. Arsic, B. Radig, and G. Rigoll. 2008. “Low-level fusion of audio and video feature for multi-modal emotion recognition.” In *Proc., 3rd Int. Conf. on Computer Vision Theory and Applications*, 145–151. New York: Science and Technology Publications.

WSDOT (Washington State Department of Transportation). 2009. “2009 bridge inspection report for bridge 08433700.” Olympia, WA: WSDOT. Accessed December 18, 2019. https://www.kingcounty.gov/transportation/SouthParkBridge/~/media/transportation/SouthParkBridge/docsIII/SPBInspectionReport.ashx.

WSDOT (Washington State Department of Transportation). 2013. *2013 bridge inspection report for bridge 0005115A*. Olympia, WA: WSDOT.

Wu, S. M., X. Y. You, H. C. Liu, and L. E. Wang. 2018. “Improving quality function deployment analysis with the cloud MULTIMOORA method.” *Int. Trans. Oper. Res.* 27 (3): 1600–1621. https://doi.org/10.1111/itor.12484.

Xiao, F. 2019. “Multi-sensor data fusion based on the belief divergence measure of evidences and the belief entropy.” *Inf. Fusion* 46 (Mar): 23–32. https://doi.org/10.1016/j.inffus.2018.04.003.

Yang, Y., and J. Eisenstein. 2013. “A log-linear model for unsupervised text normalization.” In *Proc., 2013 Conf. on Empirical Methods in Natural Language Processing*, 61–72. Stroudsburg, PA: Association for Computational Linguistics.

Yang, Y., and G. I. Webb. 2001. “Proportional k-interval discretization for naive-Bayes classifiers.” In Vol. 2167 of *Proc., European Conf. on Machine Learning*. Berlin: Springer.

Zadeh, L. A. 1965. “Fuzzy sets.” *Inf. Control* 8 (3): 338–353. https://doi.org/10.1016/S0019-9958(65)90241-X.

Zadeh, L. A. 1978. “Fuzzy sets as a basis for a theory of possibility.” *Fuzzy Sets Syst.* 1 (1): 3–28. https://doi.org/10.1016/0165-0114(78)90029-5.

Zhang, C. 2015. “Applying data fusion techniques for benthic habitat mapping and monitoring in a coral reef ecosystem.” *ISPRS J. Photogramm. Remote Sens.* 104 (Jun): 213–223. https://doi.org/10.1016/j.isprsjprs.2014.06.005.

Zhang, J., and N. M. El-Gohary. 2016. “Extending building information models semiautomatically using semantic natural language processing techniques.” *J. Comput. Civ. Eng.* 30 (5): C4016004. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000536.

Zhang, L., X. Wu, H. Zhu, and S. M. AbouRizk. 2017. “Perceiving safety risk of buildings adjacent to tunneling excavation: An information fusion approach.” *Autom. Constr.* 73 (Jan): 88–101. https://doi.org/10.1016/j.autcon.2016.09.003.

Zheng, H., and Y. Deng. 2018. “Evaluation method based on fuzzy relations between Dempster-Shafer belief structure.” *Int. J. Intell. Syst.* 33 (7): 1343–1363. https://doi.org/10.1002/int.21956.