Available online at www.sciencedirect.com

ScienceDirect

Procedia Engineering 145 (2016) 729 – 737

Procedia Engineering

www.elsevier.com/locate/procedia

International Conference on Sustainable Design, Engineering and Construction

Automatic logical inconsistency detection in the National Bridge Inventory

Zia Ud Din, M.S[a],*, Pingbo Tang, Ph.D[b]

[a]Del E. Webb School of Construction, Arizona State University, PO Box 870204, Tempe, AZ 85287-0204, USA.
[b]Del E. Webb School of Construction, Arizona State University, PO Box 870204, Tempe, AZ 85287-0204, USA.

Abstract*

Studies about the data quality of National Bridge Inventory (NBI) reveal missing, erroneous, and logically conflicting data. Existing data quality programs lack a focus on detecting the logical inconsistencies within NBI and between NBI and external data sources. For example, within NBI, the structural condition ratings of some bridges improve over a period while having no improvement activity or maintenance funds recorded in relevant attributes documented in NBI. An example of logical inconsistencies between NBI and external data sources is that some bridges are not located within 100 meters of any roads extracted from Google Map. Manual detection of such logical errors is tedious and error-prone. This paper proposes a systematical "hypothesis testing" approach for automatically detecting logical inconsistencies within NBI and between NBI and external data sources. Using this framework, the authors detected logical inconsistencies in the NBI data of two sample states for revealing suspicious data items in NBI. The results showed that about 1% of bridges were not located within 100 meters of any actual roads, and few bridges showed improvements in the structural evaluation without any reported maintenance records.

Peer-review under responsibility of the organizing committee of ICSDEC 2016

Keywords: Logical inconsistency; NBI database; Bridges condition;

1. Introduction

The estimated average annual bridge collapse rate in the United States is between 87 and 222 with an expected value of 128 [1]. These bridge failures result in fatalities, injuries, and causing enormous economic loss. These

* Corresponding author. Tel.: +1-480-558-6233; fax: +0-000-000-0000 .
E-mail address: zziauddi@asu.edu

1877-7058
Peer-review under responsibility of the organizing committee of ICSDEC 2016
doi:10.1016/j.proeng.2016.04.093

disasters continue to occur despite the fact that the conditions of all 607,380 national bridges in the United States — are inspected at regular intervals not to exceed twenty-four months with some exception with FHWA approval, and the bridge condition is recorded in the National Bridge Inventory (NBI) [2,3]. The NBI data is used to identify and prioritize which bridges are in greatest need of preventive maintenance, rehabilitation, or replacement [4] [3]. NBI data provides the basis for the Federal Highway Administration's (FHWA) decisions about bridge safety and the resource allocations made to keep bridges safe. Unfortunately, unexpected structural failures of bridges indicate that that a lot of space exists for further improving the quality and reliability of the NBI data.

Studies about the data quality of National Bridge Inventory (NBI) have found flaws, oddities, and omissions in NBI [5] despite the fact that Federal Highway Administration (FHWA) has developed some error checkers. The state department of transportations (DOTs) are also required to check the data for errors before submittals [6]. These NBI data checkers can conduct cross checks, null checks, safety checks, and follow-up checks, etc. These checks ensure the data entered according to the coding guide [6,7]. However, some logical inconsistencies like wrong spatial coordinates of bridges, are not detectable using such data checkers. Moreover, the data produced from different inspection and data processing procedures can conflict with each other even engineers strictly follow the standard NBI procedures of generating bridge inspection reports[9]. For example, at the original data level, calibration errors occur when sensors drift over time; at the data processing workflow level, uncertainties of selected data processing parameters for filtering and analyzing data can propagate across data products produced along the workflow. In many cases, uncertainty accumulations are unavoidable due to the limitation of the measuring systems, missing data due to sensor malfunction or human errors, extra data due to duplication of data collection efforts [10]. Existing data quality checkers developed by the federal and state agencies cannot ensure comprehensive scrutiny of most of the uncertainties and logical inconsistencies due to original data errors and uncertainty accumulations.

In this paper, we developed a "Hypothesis testing" framework for detecting logical inconsistencies between NBI and external data sources. Specifically, we evaluated three hypotheses: 1) bridges should be on roads; 2) bridge condition should not improve without maintenance; 3) similar bridges (based on age, material, and traffic) in similar areas should show similar conditions. This paper uses NBI data of two states as a test case and discusses bridges identified by this hypothesis-testing approach as potentially logical inconsistencies in the data. We expect that the findings through such hypothesis-testing studies will help bridge maintenance agencies to quickly filter out the potential problems in bridge inspection data and take measures to resolve potential inconsistencies and improve the resource allocation through optimal utilization of labor and resources.

2. Literature Review

Data are discrete objective facts, which are used to make informed decisions. Therefore, data quality affects the soundness of the decisions [9,10]. The first sub-section below focuses on reviewing general spatiotemporal data quality checking in various engineering domains. The second sub-section focuses on existing data quality checking methods designed in the domain of civil engineering, some of which are specifically for NBI data quality checking.

2.1. Quality checking and uncertainty analysis of spatiotemporal data

Data quality is the reliability of data to inform and evaluate decisions. Data of poor quality would not be suitable for the intended purpose [13]. Therefore, in today's world of massive electronic datasets, data quality problems can create significant economic and political inefficiencies [14]. Frequently mentioned dimensions of data quality are accuracy, reliability, consistency, precision, usefulness, timeliness, fineness, understandability, conciseness, and usefulness [15]. Researchers analyzed data quality along all these dimensions using data mining methods, which are algorithms for discovering latent knowledge, intelligible patterns, and meaningful insights from data [16]. Data mining draws on concepts from statistics, machine learning, database systems, and high-performance computing to accomplish its tasks. Another subset of data mining is spatiotemporal mining; it extracts spatial, temporal, and spatiotemporal relationships or other potentially useful patterns from datasets [17]. Spatiotemporal datasets have spatial and temporal components. The spatial component defines a universal reference space for all objects, usually geographical location on Earth's surface (indexed by latitude and longitude), whereas the temporal component is time-series data in which successive values in a series represent measures over time for a spatial location [18].