Available online at www.sciencedirect.com

ScienceDirect

Procedia Engineering 100 (2015) 1348 – 1353

Procedia Engineering

www.elsevier.com/locate/procedia

25th DAAAM International Symposium on Intelligent Manufacturing and Automation, DAAAM 2014

Using Probabilistic Models for Missing Data Prediction in Network Industries Performance Measurement Systems

Kristjan Kuhi*, Kati Kõrbe Kaare, Ott Koppel

Department of Logistics and Transport, Tallinn University of Technology, Akadeemia 15A, 12616 Tallinn, Estonia

Abstract

The vast development of information and communication technologies has created new possibilities to acquire and analyze data to take performance measurement systems to next level. Most commonly performance measurement has been known as a financial management tool. Sophisticated new technologies have made it possible to collect continuous real-time data and enabled to start designing and implementing nonfinancial performance measurement systems. Most network industries are undertakings of dominant position and therefore subjects to strict supervision. For the authorities to fulfill their regulatory functions, precise monitoring and systemized feedback on the performance of network industries is essential. The problem lies in non-complete data in terms of missing, faulty or delayed values which might lead to incorrect management decisions. The objective of this paper is to explore the use of mathematical models for missing data prediction in performance measurement systems. Applying deterministic models hide the uncertainty of the value state therefore with higher likelihood false diagnoses occur. Authors propose probabilistic models because likelihood based methods for missing data calculation are able to take into account different parameters and time aspect in a single model to convey more trustworthy estimates in performance measurement systems than traditional methods.

Peer-review under responsibility of DAAAM International Vienna

Keywords: industrial engineering; performance measurement; data collection; probabilistic graphical models

1. Introduction

Network industries can be defined as entities where the institution or its product consists of many interconnected

* Corresponding author. Tel.: + 372-51-22-953.
E-mail address: kristjan.kuhi@gmail.com

Peer-review under responsibility of DAAAM International Vienna
doi:10.1016/j.proeng.2015.01.502

nodes and where the connections among the nodes define the character of commerce in the industry [1]. A node in this context can be an institution, a unit of an institution or its product [2]. Examples of network industries are: transportation networks, postal services, electrical power networks, telecommunication networks etc.

When observed parameters of a system or a process are not delivered to processing entity (e.g. to performance measurement system, PeMS) in time, we consider the data missing. The problem of missing data afflicts a variety of application areas in network industries [3,4]. The datasets available to build models are often characterized by missing values, due to various causes such as sensor faults, problems of not reacting experiments, not recovering work situations, transferring data to digital systems [5].

PeMS are reliant on existence of data. Performance cannot be managed and controlled when the feedback cycle does not function properly or delivers faulty indications of the system state. In another hand missing data is everyday problem in statistics. Little and Rubin [6] explain that the mechanisms that lead to missing data are grouped into three distinct groups: missing completely at random (MCAR) – the absence of a data element is not associated with any other value in the data set, observed or missing; missing at random (MAR) – the absence of a values depend only on the observed values in the data set, not on the values that are missing; not missing at random (NMAR) – the absence of a value depends on the other missing values in the data set. In this paper we concentrate on the MCAR and MAR mechanisms where the absence of data does not have information value.

By applying this limitation, the methods for coping with missing values can be grouped into three main categories [6]: inference restricted to complete data (missing values are excluded), imputation-based approaches (missing values are filled in manually), and likelihood-based approaches (missing values are predicted).

The objective of this paper is to explore using mathematical models for missing data prediction in performance measurement systems. Constructing structured probabilistic models of the performance indicators (PIs) taking into consideration the surrounding indicator environment enables to find the indicator values with highest likelihood. It assists to fill gaps in data to improve the quality of the performance analysis and management decisions. In this paper authors present how to apply probabilistic graphical models to performance measurement in network industries that provides support for decision making in case of partial measurement data.

2. Hierarchical model for performance measurement

Performance measurement is the use of statistical evidence to determine progress toward specifically defined social or organizational objectives. Performance measurement describes also the feedback or information on activities with respect to meeting strategic objectives. They are used to measure and improve the efficiency and the quality of the production processes, and identify opportunities for progressive improvements in process performance. Most traditional measures overlook key non-financial PIs [7,8].

Performance is a term used in engineering, in economics and in many other areas. It can have a general meaning or a specific meaning. For the latter, and particularly for network industries, performance must be a measurable entity. Performance measurement techniques represent a key element of network industries asset management systems. Data collection for these systems is becoming feasible due to innovative technological advancements. This is essential for assessing the current and future state of specific fields and management efficiency in productivity, cost-effectiveness, environmental protection, preservation of investments and other functions.

According to literature contemporary PeMS should meet the following criteria: support strategic objectives; have an appropriate balance; have a limited number of performance measures; be easily accessible; consist of performance measures that have comprehensible specifications [9]. Other issues that should be considered selecting performance measures that can be used in evaluation includes forecast ability, clarity, usefulness, ability to diagnose problems, temporal effects and relevance [10,11].

Fig. 1 describes the conceptual model for PIs in the example of road industry. Technical parameter (TP) is measurable or observable environment characteristic. TP has value which varies over time. Uniform PIs permit an evaluation of the effects of different network design and maintenance strategies, but they can also be a basis for predicting network industries performance and for improving old and developing new prediction models. PIs are defined for different types of pavement structures and road categories. In a first step several single PIs describing the characteristic of the road pavement condition are assessed.