Expert Systems with Applications 38 (2011) 7143–7150

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Integrating data mining with KJ method to classify bridge construction defects

Ying-Mei Cheng [a,*], Sou-Sen Leu [b,1]

[a] Department of Civil Engineering, China University of Technology, 56 Hsing-Lung Road, Section 3, Taipei 116, Taiwan, ROC
[b] Department of Construction Engineering, National Taiwan University of Science and Technology, 43 Kee-Lung Road, Section 4, Taipei 10672, Taiwan, ROC

ARTICLE INFO

Keywords:
Affinity diagram
Bridge construction
Constraint-based clustering
KJ method

ABSTRACT

This paper tries to analyze common bridge construction defects, classify them into appropriate groups, and redefine them as a precautionary measure and means to improve quality in bridge construction. For this purpose, data on bridge construction since January 2007 were obtained from the evaluation report of the Public Construction Committee (PCC) of Taiwan. Bridge construction defects were classified according to their characteristics. A constraint-based clustering method and affinity diagram (KJ method) are proposed and used. This method can simultaneously treat mixed data types; moreover, it can incorporate user-specified constraints. The quality or safety issues, the unit-in-charge (Government authorities/project owners/contractor), and the properties of the defects (construction/audit/documents/others) are the sorting attributes. The constraint is avoiding empty clusters or clusters having very few objects. The results revealed five major defect classifications: safety and environment, construction site defects, supervision/control process, construction quality documents, and others.

1. Introduction

Construction defects are always a major concern in infrastructure projects. These defects directly influence the quality of the construction and may result in potential injury to the people. For example, the main 458-foot span of the 35-W interstate bridge that collapsed into the Mississippi river in Minneapolis; Part of the Big Nickel Road Bridge collapsed onto the roadway below it during construction in Canada. If constructors ensure construction quality, such disasters can be reduced. Therefore, the critical issue here is how recurring defects in such constructions can be avoided.

Graves (1993) proposed practical prevention instead of correction, the collection and communication of data graphically to track continuous improvement, and the use of a plan-do-check-act cycle and pilot projects to test and refine processes. Thus, it can be seen that an effective defect feedback system can prevent recurring defects, thereby improving construction quality. In Taiwan, the Public Construction Committee (PCC) is a body that audits important construction projects each season. The evaluation reports compiled by the PCC describe many defects in bridge construction processes; these descriptions, however, are miscellaneous and disordered. Therefore, it is very important to arrange and classify common defects, translate them into useful information, and then enable project participants to understand the critical defects so that their recurrence can be prevented.

With this background, this paper tries to integrate the KJ method (affinity diagram) and constrained *k*-prototypes (CKP) to analyze common bridge construction defects, classify them into appropriate groups, and redefine them as a precautionary measure and means to improve the quality of bridge construction. The integrated process is used to analyze the results of the PCC audit report and inspect the operation of the quality management system for public construction projects in Taiwan. Through this process, a general list of defects in bridge construction operations can be built and recurring defects can be prevented.

2. Quality management system for public construction projects in Taiwan

According to the definition of the American National Standards Institute (ANSI/ASQC, 1987), a quality system is "the organizational structure, responsibilities, procedures, processes, and resources for implementing quality management." A quality system includes the entire organization, the quality procedure, and the results of related work. In the construction industry, the system is a complex one because it involves different participants linked by reciprocating relationships, such as the owner, constructor, subcontractors, and supplier.

In order to improve public construction project quality and maintain a quality system, the PCC in Taiwan has been framing the construction quality management system Three-level Quality

* Corresponding author. Tel.: +886 2 2931 3416x2170; fax: +886 2 2934 6117.
E-mail addresses: yingmei.cheng@msa.hinet.net (Y.-M. Cheng), leuss@mail.ntust.edu.tw (S.-S. Leu).
[1] Tel.: +886 2 2733 3141x7511; fax: +886 2 2737 6606.

0957-4174/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2010.12.047

Management System (TQMS) over a period of more than 10 years. This system clearly defines the responsibilities of the owner and contractor. All the participants of a project must adhere to the TQMS, unless specified otherwise, when they carry out operations related to public works. It comprises three parts, which are discussed below:

- *Quality control (QC, first level)*: The contractor shall be in charge of quality control.
- *Quality assurance (QA, second level)*: In the TQMS, project owners perform construction quality assurance.
- *Quality audit (third level)*: Government authorities and PCC are responsible for quality audits.

The PCC used the TQMS and included the unit-in-charge of a project to improve the quality of public construction projects and implement the quality system. The final and critical process would be to audit all construction processes and results. The objective of this paper is to discuss how these defects of the audit can be arranged and classified for preventing their recurrence.

3. Methodology

Since 1996, some studies that address the problem of classification in construction management have been conducted. Holt (1996) applied a cluster analysis to classify construction contractors. Ashraf (2006) used unsupervised-learning neural networks to classify construction contractors and emphasized the promotion of the efficiency of contractor prequalification processes. Chinyio, Olomolaiye, Kometa, and Harris. (1998) proposed a need-based methodology for classifying construction clients. They reclassified the traditional classification of clients (private, public, and developer) for evaluating contractors who could satisfy the criteria for these categories in an optimal way. Dzeng (2006) presented an analytical model to identify design management packages. This model reduces the number of design interfaces between participating design firms. Tsai and Yang (2004) employed the constrained fuzzy c-mean clustering algorithm to determine bridge let projects. The majority of the above investigations have adopted the clustering method to treat classification problem issues and they have obtained good results; therefore, clustering with the KJ method has been adopted in this research to classify common bridge construction defects.

Most existing clustering algorithms can either handle mixed data types or user constraints (Bradley, Bennett, & Demiriz, 2000; Chan & Chung, 1999; Gao, Li, & Jiao, 2003; He & Xiong, 2005; Huang, 1998; Ng, 2000; Ralambondraint, 1995; Tung, Lakshmanan, & Han, 2001). Few algorithms can do both these operations well. For example, Bradley et al. (2000) proposed the constrained k-means clustering, which can avoid local solutions with empty clusters but cannot deal with categorical data. Huang (1998) presented the k-prototypes algorithm, which can cluster objects with mixed data types but cannot take any user constraints into consideration. Current clustering algorithms mainly focus on the constraints at either the cluster level or the instance level. Dan and Kamvar (2002), Wagstaff and Cardie (2001), and Davidson (2005) separately proposed constrained clustering with background knowledge. They placed importance on the pair-wise constraints at the instance level on the clustering process, such as must-link and cannot-link. However, these are not applicable to this study. The primary objective of this research is to propose the effective constraint-based clustering method CKP. The CKP algorithm can simultaneously handle user constraints and mixed data types. In addition, the algorithm is combined with the KJ method to classify the construction defects.

3.1. KJ method

The KJ method (affinity diagram) was developed by Kawakita Jiro in 1953, and it is used to organize data into useful categories or in other words, transform data into information (Decker, 2006). Plain (2007) and Lance (2006) described the characteristics of the KJ method and proposed a way to build an affinity diagram. White, Behara, and Babbar (2002) used the KJ method to organize vast amounts of qualitative data and identify customer experience patterns. Zaphiris, Ghiawadwala, and Mughal (2005) employed affinity diagrams to study a set of age-centered research-based web design guidelines. In fact, the KJ method is a typical quality technique. It adopts the bottom-up sorting process and is very useful for classifying data. It is, however, a subjective process and may consume time when the volume of data is large. To create an affinity diagram, a group of people, index cards or sticky notes, and physical space are required. The steps in the KJ method are as follows: (1) determine the theme, (2) gather data, (3) sort data into groups, (4) create header cards, and (5) draw finished diagram. In step (3), people first transfer data onto index cards or sticky notes; scatter the cards on a table or post the notes on a wall; and then according to related ideas, issues, or topics of data, arrange the cards (Plain 2007).

3.2. CKP sorting method

Huang (1998) presented the k-prototypes algorithm, which provides a straightforward way to integrate the k-means and k-modes algorithms to cluster mixed data type objects. Based on the k-prototypes algorithm, a constrained function (CF) is incorporated to the objective function of the k-prototypes by adding up and recording the distances between cluster objects after any necessary movement of objects between the clusters (Cheng & Leu, 2008). The objective function is defined as follows:

$$P(W,Q) = \sum_{l=1}^{k} \left(P_l^r + P_l^c\right) + \mathrm{CF} \tag{1}$$

The CF was obtained in a different manner, according to whether the summation of the constrained attribute value in every iteration was below the lower limit or above the upper limit. The CF varies from the movement of objects and it is given by the following equation:

$$\mathrm{CF} = \sum_{v} \Delta P^r + \Delta P^c \tag{2}$$

(1) If the constrained attribute is above the upper limit

$$\Delta P^r = -\omega_{vo}\sum_{j=1}^{p}(x_{v,j} - q_{o,j})^2 + \omega_{va}\sum_{j=1}^{p}(x_{v,j} - q_{a,j})^2 \tag{3}$$

$$\Delta P^c = -\gamma\omega_{vo}\sum_{j=p+1}^{m}\delta(x_{v,j}, q_{o,j}) + \gamma\omega_{va}\sum_{j=p+1}^{m}\delta(x_{v,j}, q_{a,j}) \tag{4}$$

(2) If the constrained attribute is below the lower limit

$$\Delta P^r = \omega_{vo}\sum_{j=1}^{p}(x_{v,j} - q_{o,j})^2 - \omega_{vs}\sum_{j=1}^{p}(x_{v,j} - q_{s,j})^2 \tag{5}$$

$$\Delta P^c = \gamma\omega_{vo}\sum_{j=p+1}^{m}\delta(x_{v,j}, q_{o,j}) - \gamma\omega_{vs}\sum_{j=p+1}^{m}\delta(x_{v,j}, q_{s,j}) \tag{6}$$

Here, $x_{v,j}$ is the j attribute of a moving object v; $q_{o,j}$, center of the j attribute of the original cluster (the cluster which is below the lower limit or above the upper limit); $q_{a,j}$, center of the j attribute of the acceptable cluster (the cluster which accept the object from the original cluster); and $q_{s,j}$, center of the j attribute of the supply cluster (the cluster which supply the object to the original cluster.