Automation in Construction 19 (2010) 791–797

Contents lists available at ScienceDirect

Automation in Construction

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

Predicting construction crew productivity by using Self Organizing Maps

Emel Laptali Oral [a,*], Mustafa Oral [b]

[a] Çukurova Üniversitesi, Mühendislik Mimarlık Fakültesi, İnşaat Mühendisliği Bölümü, Balcalı, Adana, Turkey
[b] Çukurova Üniversitesi, Mühendislik Mimarlık Fakültesi, Bilgisayar Mühendisliği Bölümü, Balcalı, Adana, Turkey

ARTICLE INFO

Article history:
Accepted 5 May 2010

Keywords:
Productivity prediction
Artificial neural networks
Self Organizing Maps
Construction crews
Formwork
Reinforcement
Concrete
Work study

ABSTRACT

A Self Organizing Map (SOM), is a machine learning method that represents high-dimensional data in low-dimensional form without losing topological relations of the data. After an unsupervised learning process, it organizes the data on the basis on similarity. In the current study, a SOM based algorithm has been developed which not only produces 2-D maps to analyze the relationship between various factors and crew productivity, but also predicts productivity under given conditions. Validation of the model has been achieved both by using artificial data set and data from 144 concrete pouring, 101 formwork and 101 reinforcement crews. The results show that maps which are produced by the model are satisfactory in clustering the data and prediction performance of the model is superior to similar artificial neural network models.

1. Introduction

Various statistical methods have been used to analyze the effect of project, management, environment and crew related factors on construction labour productivity [1–13]. Meanwhile, due to the complex variability of crew productivity data and insufficiency of statistical diagnostics in dealing with the relationships, some researchers have focused on artificial neural network applications based on supervised learning process. Performance of these applications has generally been compared to multiple regression models and strengths of the former in both modelling ease, analysis flexibility and prediction results have been significant even when multiple regression results demonstrate good correlations [14–24]. However, weaknesses of the supervised learning process which the output vector has to be known for training have been as well pointed out [21,25]. In parallel, literature shows that Self Organizing Maps (SOM) overcome the disadvantages of both statistical methods and neural network applications which are based on supervised learning, by using unsupervised learning [25–28]. However, most of the current applications from medicine to economics focus only on the use of SOM in analyzing the relationship between two or more variables by using two-dimensional visualization. Few researchers like Hwa and Miklas [29], Du et al. [30] and Mochnache et al. [31] used SOM for prediction of heavy metal removal performance, oil temperature of transformers and thermal ageing of transformer oil, respectively. This research, thus, focus on developing a SOM based model in order to both analyze the relationship between various site, management, crew, and environment related factors and crew productivity and also make crew productivity prediction under given conditions.

2. Problem definition and data collection

The underlying problem of the current research has been to develop a novel computer based system which will not only ease the work of site management in documentation and monitoring of construction crew productivity, but also provide an environment for analysis and prediction of construction crew productivity under observed conditions.

To achieve this, standard time study sheets have been developed to systemise the data gathering procedure for concrete, formwork, reinforcement, masonry, plastering, painting and floor covering crews [32]. Standard time sheets include sections related with work and crew characteristics together with site, management and environmental factors that may affect crew productivity. Labour related factors are age, education, experience, working hours, payment method, absenteeism and crew size. Work related factors are location of the site, location of the work on site, the type and the size of the material used and the weather conditions. Site management factors are site congestion, transport distances, and, availability of the; crew, machinery, materials, equipments and site management. This paper will focus on prediction results for ready mixed concrete, formwork and reinforcement crews with work definitions given in Table 1 below.

Related data were collected between September 2006 and September 2008 from various randomly selected construction sites in Turkey. Pilot studies with 30 crews from each activity group were first undertaken and depending on the variation of data (using Eq. (1)), the minimum amount of crews that had to be observed was

* Corresponding author.
E-mail addresses: eoral@cu.edu.tr (E.L. Oral), moral@cu.edu.tr (M. Oral).

0926-5805/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.autcon.2010.05.001

Table 1
Definition of the work items.

Activity group	Definition of the work item
Ready mixed concrete	Pump from the transmixer
	Vibrate the concrete
	Level the concrete
	Protect the concrete from hot/cold
	Water the concrete
	Take samples for the quality control of the concrete
Formwork	Carry the scaffolding
	Erect the scaffolding
	Grease the formwork
	Dismantle the formwork
	Clean the formwork
	Dismantle the scaffolding
Reinforcement	Unload the steel from the trucks
	Carry the steel within the site
	Cut the steel
	Bend the steel
	Lay down the steel

determined. After data collection, outliers were eliminated by using Box and Whiskas method [33]. The number of observations is presented in Table 2.

$$N'\left[A\frac{\sqrt{N\sum_{i=1}^{n}X_i^2-\left(\sum_{i=1}^{n}X_i\right)^2}}{\sum_{i=1}^{n}X_i}\right]^2 \quad (1)$$

where:

- N' Required number of observations within the targeted confidence interval.
- A 40 for 95% confidence level, $A=20$ for 90% confidence level.
- N Number of observations during the pilot study.
- X_i Unit output of the related labour (crew) during the ith observation.
- n Number of observations during the pilot study.

After outlier elimination, number of crews for ready mixed concrete fulfilled the requirements for 95% confidence level and reinforcement and formwork fulfilled the requirements for 90% confidence level. Mean productivity values were then calculated for each activity group and normality of the data were tested by determining the skewness (to be between ± 3), and kurtosis (to be between ± 10) coefficients for productivity values for each activity group [34]. The strong positive skewness (skewness coefficients being very close to 3) and leptokurtic kurtosis (kurtosis coefficients being very close to 10) showed that the observations were spread in a wide range and not suitable for statistical methods based on normality assumptions (Table 3).

3. Self Organizing Maps (SOM)

"A self-organizing map is a type of artificial neural network that is trained using unsupervised learning to produce a low-dimensional (typically two-dimensional), discretized representation of the input space of the training samples, called a map. The map seeks to preserve the topological properties of the input space" [35].

SOM was first developed by Kohonen [36] and used for speech recognition. Since then, SOM has been theoretically developed continuously and has been widely used in various areas like economics, biology, medicine and engineering. Over 5000 publications have been reported in the literature. SOM combines two important aspects, which are; reducing the amount of data by clustering, and organizing the data on a low-dimensional display. Ability of easy two-dimensional visualization is SOM's superiority over other data analysis tools. Output vector is not required to be known, i.e. the system does not use pairs consisting of an input and the desired output for training but instead uses the input and output patterns and locates remarkable patterns, regularities or clusters among them. In other terms, SOM discovers groups of similar instances, instead of requiring a predefined classification. So if the causal relation between the input and output observations has a complex variability, it is often easier to bridge the gap using unsupervised learning instead of supervised learning [37].

A SOM network has two layers. First one, which is 'Input Layer', has nodes equal to the number of parameters. Each input node is connected to the nodes in the 'Output Layer' with a weight constant. Number of nodes in the 'Output Layer' should be small enough to ensure fast computation, and large enough to display any relation between the parameters. SOM networks learn to detect regularities and correlations in their input and adapt their future responses to that input accordingly. Fig. 1 shows an example, which adapts a 4-dimensional input vector into a 6×6 dimensional map. Adaptation process involves organizing unknown data into groups of similar patterns, according to a similarity criterion (e.g. Euclidean distance). Thus, when the process is complete, each node in the 'Output Layer' has a topological position where similar clusters position close to each other.

Table 2
Required and actual numbers of observations.

Activity group	Min. no. of crews required		No. of observed crews	No. of crews after outlier elimination
	95% confidence level	90% confidence level		
Ready mixed concrete	54	27	175	144
Reinforcement	118	59	184	101
Formwork	120	60	162	101

Table 3
Distribution characteristics of productivity values.

	Concrete (h/m^3)	Reinforcement (h/kg)	Formwork (h/m^2)
Mean productivity	0.29	0.034	0.87
Standard deviation	0.22	0.028	0.85
Coefficient of variation	0.76	0.82	0.97
Skewness coefficient	2.88	2.35	2.82
Kurtosis coefficient	9.66	6.1	9.57

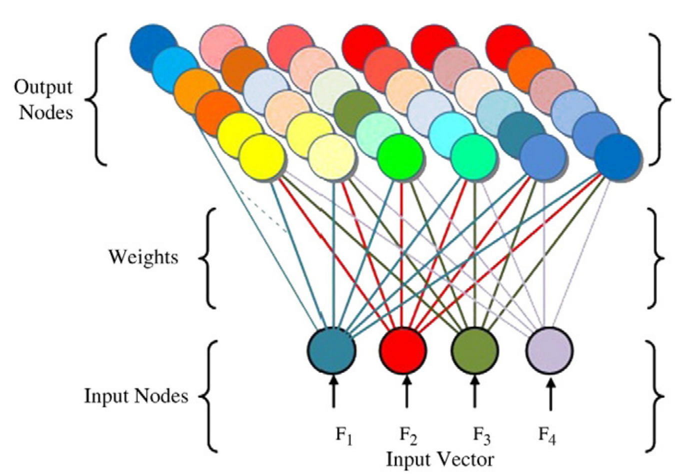

Fig. 1. SOM structure.