5th International/11th Construction Specialty Conference
5e International/11e Conférence spécialisée sur la construction

Vancouver, British Columbia
June 8 to June 10, 2015 / *8 juin au 10 juin 2015*

CONSTRUCTION PRODUCTIVITY MODEL USING FUZZY APPROACH

Emad Elwakil[1, 4], Tarek Zayed[2], Tarek Attia[3]
[1] Building Construction Management Department, Purdue University, West Lafayette, IN 47907, United States
[2] Department of Building, Civil & Environmental Engineering, Concordia University Montreal, Canada
[3] Institute of Construction Engineering and Management, Housing and Building National Research Center, Egypt
[4] eelwakil@purdue.edu

Abstract: Productivity is one of the most important elements to manage construction projects especially with regards to the prediction of the activities' durations. Uncertainty is an entrenched characteristic of most construction projects. Most research works in simulating construction productivity have focused predominantly on modeling and have neglected to study the effect of subjective variables on productivity of construction process. The unique nature of construction projects and uncertainty of the construction processes lead to a need of new generation of models that utilizes the historical data. The presented research develops, using Fuzzy approach, a model to utilize, analyze, extract and find the hidden patterns of the project data sets to predict the construction process productivity. The engine depends on finding the relation between quantitative and qualitative variables, which affect the construction processes, and productivity. The methodology of this research consists of six steps: (1) Investigate the factors affecting the productivity (2) select the critical factors that affect the productivity; (3) build Fuzzy sets; (4) generate Fuzzy rules and models; (5) build Fuzzy knowledge base; and (6) validate the effectiveness of the built model to predict the construction process productivity. The developed model is validated and verified using case study with sound and satisfactory results, 90.65 % average validity percent. The developed research/engine benefits both researchers and practitioners because it provides robust model for construction processes and a tool to predict the productivity of construction processes.

1 INTRODUCTION

Productivity is one of the most important elements to manage construction projects especially with regards to the prediction of the activities' durations. Productivity is defined as "the ratio of output of required quality to the inputs" for a specific construction process (Al-Zwainy, et. al. 2013). Several studies applied many statistical methods in the construction management; regression, probability functions, stochastic techniques, and mathematical learning curves for simulation and/or optimization (Bee Hua, 2008).

Raw (i.e. dirty) data can cause confusion for the mining and modeling procedure that leads to unreliable output (Han and Kamber 2006). If users believe the data are dirty, they are unlikely to trust the results of any data mining that has been applied to it. Most research works in simulating construction processes have focused predominantly on modeling and neglected to prepare the data before mining process. Before dealing with the system building, the data problem should be identified. There are major two problems of data, which are missing data and outliers. The simple way that most research works deal with is to remove incomplete data set and outliers. Knowledge Discovery in Database (KDD) is the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data (Fayyad et. al 1996). Data mining is used to extract hidden knowledge from a data set that is not obtained by traditional methods as statistical analysis. The KDD is an interdisciplinary field involving concepts from machine learning, database query, statistics, mathematics and visualization (Anand et al. 1998). Most modern KDD tools have focused almost exclusively on building models (Cox and Wills 1997). A 60% of time goes into preparing data for mining; however, actual mining step typically constitutes about 10% of the overall effort (Cabena et al. 1998). Knowledge discovery in databases (KDD) and data mining consist of a series of steps including domain and data understanding, data preparation, data mining, and finally, pattern evaluation and deployment (Chapman et al. 1999). The KDD has been applied in the area of construction and facility management in recent research in response to the explosive growth of computerized historical databases (Buchheit et al. 2000; Soibelman and Kim, 2002). Chau et al. (2002) developed an application of data warehouse and decision support system to provide the information about and insight of existing data.

The previous studies show a lack of using the historical data and experts to build models/knowledge bases that help to predict the construction productivity.

2 RESEARCH OBJECTIVES

The overall objective of this research is to build a construction productivity model. To achieve this objective, the following sub-objectives are as follow:

1. Develop a data mining engine to utilize, analyze, extract, and model the hidden patterns from the project data.
2. Integrate the enormous amount of historical data and knowledge base design in order to predict construction productivity.
3. Verify and validate the proposed framework using a case study.

3 BACKGROUND

Construction management research used the fuzzy set theory, fuzzy logic, and hybrid fuzzy techniques (Chan, Chan, & Yeung, 2009). Therefore, many studies went through to combine the fuzzy set theory/ fuzzy logic with artificial intelligence systems as well as Gas, Anns, and PSO in order to create a hybrid model to implement model parameter optimization (Bee Hua, 2008).Several studies investigated the factors that affect the construction labors productivity variations; the National Electrical Contractors Association (NECA) studied the effects of humidity and temperature on labor productivity (Sonmez & Rowings, 1998). Another study developed a model using a methodology based on the regression and neural network modeling techniques to evaluate many factors on productivity for concrete pouring, formwork, and concrete finishing tasks from task to task (Sonmez & Rowings, 1998).

The theory behind learning curve models reveled that as the production quantity of any product doubles, the unit or average cost (hours, man-hours, dollars, etc.) will decreases by a fixed percentage of the previous unit or average rate (Thomas,et. al. , 1986). A number of mathematical models were used to describe and predict the learning curve, including the straight line power model, cubic power model exponential model, and piecewise model (Thomas, Mathews, and Ward, 1986). Straight line power model is the most commonly used model in construction productivity, moreover, it provides the most reliable prediction of future performance (Everett and Farghal, 1994). Hildreth (2012) presented a case study of his construction students where they participated in virtual construction operation. The experiment resulted in excellent knowledge regarding learning curve theory and its application within construction industry.

Malyusz and Pem (2014) used mathematical algorithms to evaluate the predictive capabilities of various learning curve models and data presentation methods for labor intensive construction operations. The algorithms gave sequential predictions for future performance of construction activities.

Situation-based simulation models is a recently developed technique that is used to model the triggering situations in construction to predict productivity. Choy and Ruwanpura (2006) developed an application model that uses “root causes of productivity loss” as the situations. This model allows users to develop models based on the interactions of various situations and work types. A study by (Al-Zwainy et al., 2013 developed a model using Multivariable Linear Regression technique (MLR) to estimate construction productivity for marble finishing works of floors.

4 RESEARCH METHODOLOGY

This research consists of five phases that provides detailed explanation of the research methodology. A comprehensive literature review phase has been conducted that includes the state of the art review of construction productivity, data mining and knowledge discovery, as well as fuzzy knowledge base building and Design. The second section pertains to system development and implementation procedures that includes the following three stages: Variable Selection, Fuzzy Sets, Fuzzy Rule Induction, and Fuzzy Knowledge Base. The third phase focuses on data collection, which includes a case study to verify and validate the developed system. The fourth phase denotes verification and validation of the developed system using a case study. The final phase of this research propounds conclusions and recommendations for future work. What follows is a detailed explanation of the aforementioned five-study section and their sub sections.

5 PRODUCTIVITY DATA MINING ENGINE DEVELOPMENT

The productivity Data mining engine make the variable selection to build the knowledge base through Fazzification and fuzzy induction rules. The productivity data mining engine consists of five steps as follows:

5.1 Variables selection

In this step, the variables that affect the task productivity rate are selected using Fuzzy average method. The fuzzy curve will be built using Equation 1 and the variables will be ranked using equation 2.

I. Build the Fuzzy curve

For every input and output points (Xi,Yj) the equation 1 will be applied (Lin and Cunningham III 1995). If there is a completely random relationship between the input(s) and output(s), then, the fuzzy curve is flat and vice versa.

$$\mathrm{C_i}(\mathrm{x_i}) = \frac{\sum_{j=1}^{m} y^j v_{ij}(x_i)}{\sum_{j=1}^{m} v_{ij}(x_i)} \quad (1)$$

Where:

Ci(xi): Curve points

II. Ranking the variables

Mean Square Error (MSE) will be used to rank the variables using Equation 2. The MSE value shows the significant; if the value is small that mean a high significance and vice versa.

$$MSEci = \frac{1}{M} \sum_{k=1}^{M} (ci\,(xi, k) - yk)^2 \quad (2)$$

5.2 Fuzzy sets (Fuzzification)

Crisp quantities are converted into fuzzy sets in this step using the Artificial Neural Network (ANN) technique. The ANN will be built from modeling data sets of the selected variables. The training will simulate the relationship between the coordinate locations and membership values. After the net is trained, its validity and efficiency can be checked using the testing data. Then it is ready and can be used to determine the membership value of input data set in the different regions.

5.3 Fuzzy rules induction

Rule induction creates a knowledge base of fuzzy if-then rules by modeling the embedded patterns in the data. This is a supervised knowledge discovery that use Fuzzy rule induction algorithm. The functional relationships between the independent and dependent variables are expressed as a set of fuzzy if-then rules. The inputs of this step are Fuzzy sets that will be isolated to data point mapping with the highest membership degree that could happen if there are two data sets in the same Fuzzy set, the higher degree would be selected. The final rules are selected from a pool of data pairs. The degree of effectiveness (E) for each rule will be computed. It selects the most effective rules amongst the candidate rules using Equation 3 (Cox, 2005).

$$\mathrm{E(ri)} = \mu\mathrm{x(v1)} \times \mu\mathrm{y(v2)} \times \mu\mathrm{z(v3)} \qquad (3)$$

5.4 Fuzzy knowledge base (FKB)

The collection of all the previous steps are forming Fuzzy knowledge base, which includes variables, fuzzy sets, and fuzzy rules. This collection of FKBs may form a solution to a complex system. FKBs are containers where they store the data definitions (variables) and the required rules to solve a set of outcomes associated with the system.

5.5 Validation

Prediction effectiveness of the developed engine will tested in this step using a mathematical validation. The average invalidity percent (AIP) and the Average validity Percent (AVP) will be calculated using Equation 4 and 5 respectively. If the AIP value is closer to 0.0, the model is sound and a value closer to 100 shows that the model is not appropriate to predict the productivity (Zayed and Halpin, 2005).Similarly, the Root Mean Square Error (RMSE) is estimated using Equation 6. If the value of the RMSE is close to 0, the model is sound and vice versa. In addition, the Mean Absolute Error (MAE) is defined as shown in Equation 7. The MAE value varies from 0 to infinity. However, the MAE should be close to zero for sound results (Dikmen et al. 2005).

$$AIP = \left(\sum_{i=1}^{n} \left|1 - \left({E_i}/{C_i}\right)\right|\right) * 100/n \qquad (4)$$

$$AVP = 100 - AIP \qquad (5)$$

$$RMSE = \frac{\sqrt{\sum_{i=1}^{n}(C_i - E_i)^{2}}}{n} \qquad (6)$$

$$MAE = \frac{\sum_{i=1}^{n} |C_i - E_i|}{n} \quad (7)$$

Where:

AIP: Average Invalidity Percent

AVP: Average Validity Percent

$RMSE$: Root Mean Squared Error

MAE: Mean Absolute Error

E_i : Estimated Value

C_i : Actual Value

6 ANALYSIS OF DATA MINING ENGINE IMPLEMENTATION TO A CASE STUDY

In this study, data were collected through both on-site observation and digital camera monitoring system for Engineering, Computer Science and Visual Arts complex of Concordia University (Khan, 2005). The implementation to a case study will concern the productivity of concrete pouring operation. The factors that affect the productivity are shown in Table 1. A sample of case study data is shown in Table 2

Variables selection

In this step, the variables that affect output variable (work task productivity) are selected. Two methods are used to select the variables that affect the productivity. The first is sorting factors according to Mean Square Error (MSE) as shown in Table 3. The second method is using Fuzzy curve. As shown in Figure 1 and Figure 2. The MSE method will be used to validate the Fuzzy curve method and how it selects and ranks critical variables. The ANN method is used to compare the ranking and selection efficiency using the selected variables as shown in the following sections.

Table 1: Concrete Pouring Process Variables

No.	Variables	Description
1	Temperature ºC	Average of eight working hours of the day
2	Humidity (%)	Average of eight working hours of the day
3	Precipitation	Incorporated in terms of four numerical values as follows: No precipitation = 0, Light rain = 1, Rain = 2, and Snow = 3
4	Wind Speed (km/h)	Average of eight working hours of the day
5	Floor Height	The floor number
6	Work Type	Two types of activities will be considered as follows: Slabs = 1 and walls = 2
7	Gang Size (workers)	Number of persons in the gang
8	Labor Percent (%)	The percentage of the labor (non- skilled workers) in the gang
9	Time (min)	Work Task Duration

Table 2: A Sample of Concrete Pouring Process Variables Data

Temperature ºC	Humidity	Precipitation	Wind speed (K/h)	Gang size (workers)	Labor %	Floor level	Method	P (m3/h)
-4	35	0	13.5	12	57	4	1	4.5
-13.5	70	1	8	7	55	2	2	8.5
8	83	1	6	10	56	8	1	6.4

Variables selection using the ANN method

The ANN method is used to validate the Fuzzy Average method of selecting and ranking critical variables. Ranking of input variables is required in order to determine the relative importance of each variable and those that have the most effect on the task duration. The contribution percentages are derived from analysis of weights of the trained neural network. The higher the number, the more that variable is contributing to the classification and/or prediction. Obviously, if a certain variable is highly correlated, the variable will have a high contribution percentage. Table 4 shows the contribution percentage (relative significance) of eight variables. By comparing the results of Fuzzy Average Method and ANN methods, it is clear that four out of eight variables are similar in the top of each ranking and the other variables have different ranking.

Table 3. Variables Ranking According to Mean Square Error (MSE)

Variables	**MSE**	**Rank**
Temperature(ºC)	6.4	2
Humidity (%)	8.5	4
Precipitation	8.9	5
Wind speed (K/h)	10.72	7
Floor level	10.5	6
Method	13.5	8
Gang Size(workers)	3.25	1
Labor Percent (%)	6.5	3

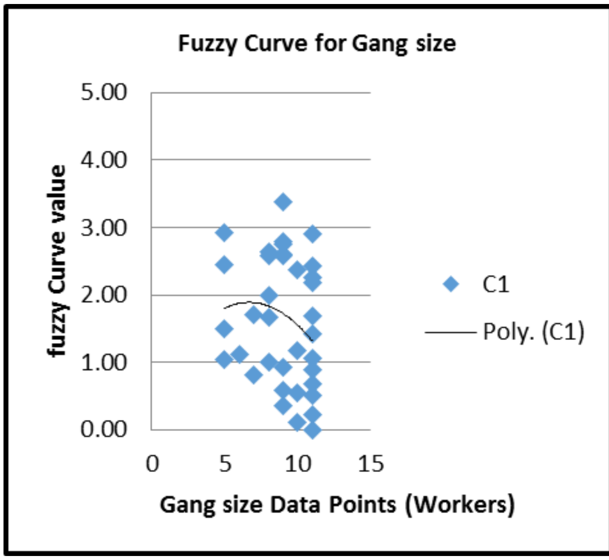

Figure 1 Fuzzy curve for Gang Size

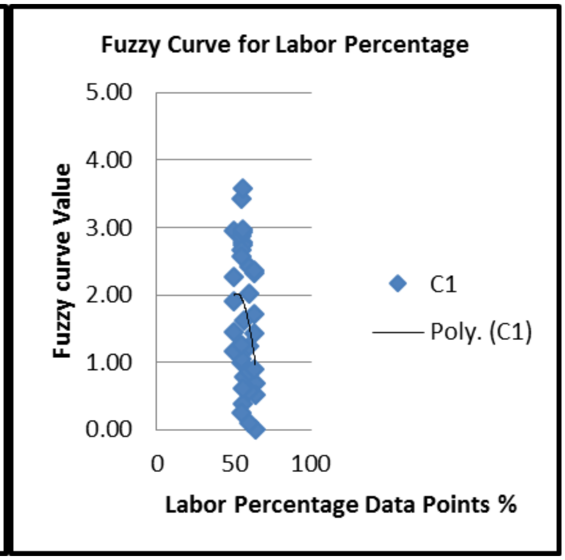

Figure 2 Fuzzy curve for Labor Percentage

Table 4. Variables Ranking Using ANN

Rank	Variables
2	Temperature
4	Humidity
8	Precipitation
5	Wind speed
7	Floor level
6	Method
1	Gang Size
3	Labor Percentage

Fuzzy sets

The Fuzzy clustered points are used to train and build the Neural Network. As shown in Table 5, a sample of Fuzzy membership values for data shown in Table 2. The ANN is used to predict and form the Fuzzy sets. The Fuzzification process involves assigning membership values for the given crisp quantities. Membership's values are assigned using the Neural Network technique in order to model the relation between the fuzzy membership functions and productivity.

Table 5 .A Sample of Fuzzy Membership Values

Cluster 1	Cluster 2	Cluster 3
0.4686	0.3416	0.1898
0.4438	0.314	0.2422
0.4764	0.3072	0.2164

Fuzzy rule induction

Rule induction creates a knowledge base of fuzzy if-then rules by modeling the embedded patterns in the data. This step aims at finding the functional relationships between variables and task duration defined as a set of “if-then” rules. Fuzzy Rule Induction Algorithm is used. A sample of Fuzzy rules is shown in Table 6. As the final rules are selected using only the highest degree of effectiveness (E).

Fuzzy knowledge base

Fuzzy knowledge base includes all the previous steps, i.e. the selected variables, fuzzy sets, and fuzzy rules. The FKB is a representation of a particular model of each process that means the FKB of concrete pouring process is separate from the loading process.

Table 6: A Sample of Fuzzy Rules and Fuzzy Knowledge Base Models

Rule 1	If C1 is	0.4686	and C2 is	0.3416	and C3 is	0.1898	Then Time is	4.5 (m3/h)
Rule 2	If C1 is	0.4438	and C2 is	0.314	and C3 is	0.2422	Then Time is	8.5 (m3/h)
Rule 3	If C1 is	0.4764	and C2 is	0.3072	and C3 is	0.2164	Then Time is	6.4 (m3/h)

Fuzzy knowledge base validation

The goal of this step is to test the engine prediction effectiveness. To test the effectiveness of the engine’s prediction, a validation data set is embedded into the developed engine to compare its results with actual data. As shown in Equation 4 and 5, the average validity percent (AVP) and the average invalidity percent (AIP) are used. The developed Knowledge base is validated by comparing the predicted results with the actual values for productivity using the validation data set. The results show that the average validity percent is 90.65 %, the RMSE is 1.5, and the MAE is 3. Therefore, the developed fuzzy knowledge base is acceptable and robust to predict the productivity of the work task.

7 CONCLUSIONS

The current research presents a construction productivity data mining engine to prepare, utilize, analyze and extract the hidden patterns from the project data to predict the work task productivity. The engine depends on preparing the historical data to be modeled and finding the relation between quantitative and qualitative variables, which affect the construction processes, and work task productivity. The developed data mining engine consists of the following steps: (1) select the factors that affect the construction process; (2) generate Fuzzy sets; (3) define Fuzzy rule and models; (4) build Fuzzy knowledge base; and (5) validate the effectiveness of the built knowledge base to predict the work task durations. The presented research shows the significant effect of qualitative and quantitative variables on the task durations. It shows the efficiency of Fuzzy approach to model the hidden patterns of construction project data. The developed engine is validated and verified using case study with sound and satisfactory results, 90.65 % average validity percent. The developed research/engine benefits both researchers and practitioners because it provides robust knowledge base for construction processes and a tool to predict the related task productivity for construction activities.

References

Al-Zwainy, F. M. S., Abdulmajeed, M. H., & Aljumaily, H. S. M. 2013. Using Multivariable Linear Regression Technique for Modeling Productivity Construction in Iraq. Open Journal of Civil Engineering, 03(03), 127–135. doi:10.4236/ojce.2013.33015

Anand, S. S., Patterson, D., Hughes, J. G., Bell, D.A. 1998.Discovering case knowledge using data mining. Proceedings of the Second Pacific-Asia Conference in Knowledge Discovery and Data Mining (PAKDD), Melbourne, Australia, 1998; pp.25-35

Arditi, D., & Mochtar, K. 2000. Trends in productivity improvement in the US construction industry. Construction Management and Economics, 18(1), 15–27. doi:10.1080/014461900370915

Bee Hua, G. 2008. The state of applications of quantitative analysis techniques to construction economics and management (1983 to 2006). Construction Management and Economics, 26(5), 485–497. doi:10.1080/01446190801998716

Buchheit, R., J. H. Garrett Jr., S. R. Lee, R. Brahme.2000 .A Knowledge Discovery Framework for Civil Infrastructure: A Case Study of the Intelligent Workplace. Engineering with Computers Journal, 16, 2000, pp.264-274.

Cabena, P., Hadjinian, P., Stadler, R., Verhees, J., & Zanasi, A.1998. Discovering data mining from concept to implementation 1998; NJ: Prentice Hall.

Chapman, P., Clinton, J., Khabaza, T., Reinartz, T., Wirth R.2008. The CRISP-DM Process Model Discussion. 03/1999 http://www.spss.it/download/pub-paper.pdf (16.06.2008; 11.45)

Chan, A. P., Chan, D. W., & Yeung, J. F. 2009. Overview of the application of “fuzzy techniques” in construction management research. Journal of Construction Engineering and Management, 135(11), 1241–1252. Retrieved from http://ascelibrary.org/doi/abs/10.1061/(ASCE)CO.1943-7862.0000099

Chau Ying Cao, K.W., M. Anson, Jianping Zhang .2002.Application of data warehouse and Decision Support System in construction management. Automation in Construction, 2002;12(2), pp.213-224.

Choy, E., & Ruwanpura, J. Y. 2006. Predicting construction productivity using situation-based simulation models. Canadian Journal of Civil Engineering, 33(12), 1585-1600.

Cox, Eick, Wills, Brachman. 1997. Visual Data Mining: Recognizing Telephone Calling Fraud. Data Mining and Knowledge Discovery, Vol. 1, 1997; pp. 225-231.

Deogun, J. Spaulding, W. Shuart, B., Li, D. 2004.Towards Missing Data Imputation: A study of fuzzy K-means Clustering Method, In Rough Sets and Current Trends in Computing .Lecture Notes in Computer Science 3066, Springer-Verlag, 2004, pp.573-579.

Everett, J. G. and S. Farghal, 1994. “Learning Curve Predictors for Construction Field Operations,” Journal of Construction Engineering and Management, 120(3), 603-616

Fayyad, U, Piatetsky-Shapiro, G. Smyth, P.1996. From data mining to knowledge discovery: an overview, Advances in knowledge discovery and data mining. American Association for Artificial Intelligence, Menlo Park, CA, 1996.

Jiawei, Han and Micheline, Kamber . 2006. Data Mining Concepts and Techniques. Second Edition, Elsevier 2006; Inc. San Francisco, CA.

Khan, Zafar Ullah .2005.Modeling and parameter ranking of construction labor productivity. Master’s thesis, Concordia University.

Kivrak S, Arslan G, Dikmen I, Birgonul, MT.2008. Capturing knowledge in construction projects: knowledge platform for contractors. 2008, Journal of engineering Management, volume 24:pp 87–95.

Malyusz, L., & Pem, A. 2014. “Predicting Future Performance by Learning Curves”. Procedia-Social and Behavioral Sciences, 119, 368-376.

Soibelman, L. and Hyunjoo, K. 2002.Data Preparation Process for Construction Knowledge Generation through Knowledge Discovery in Databases. Journal of Computing in Civil Engineering, American Society of Civil Engineers (ASCE), January 2002, Volume 16, Number 1 pp 39-48.

Soibelman L, Liu L, and Wu, J. 2004.Data Fusion and Modeling for Construction Management Knowledge Discovery 10th International Conference on Computing in Civil and Building Engineering, June, 2004, Weimar, Germany.

Sonmez, R., & Rowings, J. E. 1998. Construction labor productivity modeling with neural networks. Journal of Construction Engineering and Management, 124(6), 498–504. Retrieved from http://ascelibrary.org/doi/abs/10.1061/(ASCE)0733-9364(1998)124:6(498)

Thomas, H. R., C. T. Mathews, and J. G. Ward, 1986. “Learning Curve Models of Construction Productivity,” Journal of Construction Engineering and Management, ASCE, 112(2), 245-258.
Zayed, T. M., & Halpin, D. W. (2005). Pile construction productivity assessment. Journal of Construction Engineering and Management conference, 131, 705-714.