ESTIMATING PROJECT S-CURVE BASED ON PROJECT ATTRIBUTES AND CONDITIONS

L.-C. CHAO*

Department of Construction Engineering, National Kaohsiung First University of Science and Technology, Taiwan, ROC

ABSTRACT

Project control in construction commonly uses the S-curve that represents a project's cumulative progress overall, so obtaining a reasonable S-curve estimate has always been deemed important. The empirical method of basing the estimate on past projects and using a mathematical formula for generalizing progress as a function of time to simplify estimation can be a useful approach. A previous research proposed a cubic polynomial function for generalizing S-curves as well as a neural network model for estimating the polynomial's two parameters for producing S-curves. The present study aims to develop an improved model by changing the input and output factors so as to increase accuracy of progress prediction. Since the inflection point is an S-curve's key geometric feature and the position and slope of the inflection point can replace the polynomial function's two parameters, they are used as model outputs and the range and meaning of their values suitable for representing project progress are first established. Next, because the position of the inflection point indicates where project progress peaks and its slope indicates the extent of concentration of project progress, they are presumably connected with project schedule performance, and so they should be influenced by project conditions. Therefore, in addition to the project attributes used previously, two factors relating to project conditions, i.e. degree of project difficulty and competence of project participants are also used as model inputs. Then, data on the actual progress, project attributes, and project conditions for recently completed projects in Taiwan will be collected for building fuzzy inference systems that establish the input-output relationships using the fuzzy clustering method along with hybrid training. The accuracy of the new model in progress prediction will be tested and compared with that of the previous model. It is expected to achieve better performance.

Keywords: construction project, S-curve, polynomial, inflection point, fuzzy inference system.

1. INTRODUCTION

The S-curve represents the cumulative progress of a construction project from start to finish and its slope indicates the progress per unit of time, which is small at the beginning, gradually increases to the maximum at the inflection point, and then decreases towards the end. The shape shows changes

* Corresponding author and presenter: Email: chaolc@nkfust.edu.tw

in progress from being slow to fast and slow again, which is due to the distribution of work peaking at certain stage when the work is relatively concentrated and generally applies to any project that consists of activities whose times overlap to some extent. Using single numbers to represent the overall progress at each time point, the S-curve is simple and easy to comprehend and has long been widely used in construction as a tool for project schedule control. Therefore, for both owners and contractors, obtaining a reasonable S-curve estimate has always been important.

The traditional approach to estimating an S-curve is based on a schedule of planned activity times and progress calculation using the percent weight of each activity in the project and the percent complete of each activity at each time point. However, because the actual times of the activities comprising a project are subject to the influences of many uncertain factors, a schedule based S-curve estimate also involves much uncertainty. Alternatively, the empirical method of basing the estimate on actual progress data of past similar projects and using a mathematical formula for generalizing percent progress as a function of percent time to simplify data processing can be useful. Over the years, various formulas and ways of estimating S-curves have been studied, e.g. Kenley and Wilson (1986), Skitmore (1992), Kaka (1999), Blyth and Kaka (2006). A previous research by Chao and Chien (2009) proposed a cubic polynomial function for generalizing S-curves as well as a neural network model for estimating the polynomial's two parameters for producing S-curves. The present study aims to develop an improved model by changing its input and output factors so as to increase the accuracy of progress prediction.

2. POLYNOMIAL S-CURVE FORMULA AND PREVIOUS S-CURVE MODEL

Chao and Chien (2009) proposed the following S-curve formula with two parameters *a*, *b*:

$$y = ax^3 + bx^2 + (1 - a - b)x \tag{1}$$

where *y* and *x* denotes standardized (percent) cumulative progress and standardized (percent) time.

Equation (1) can meet an S-curve's boundary conditions of ($x = 0$, $y = 0$) and ($x = 1$, $y = 1$) and has a more succinct form than other formulas, hence it is more convenient when used in calculating progress. By using proper values of *a*, *b*, a suitable S-curve with an inflection point connecting two arcs within the boundaries can be obtained and using different values of *a*, *b* can produce different S-curves. Fitting equation (1) to actual project progress data is achieved by using the least squared error method for solving *a* and *b* for an S-curve closest to all data points; the solutions are given in Chao and Chien (2009). Root of mean squared error (*RMSE*) is used to measure the accuracy of an S-curve formula in fitting to progress data and for model performance evaluation, as defined next:

$$RMSE = \sqrt{\frac{\sum_{i-1}^{d} (\hat{y}_i - y_i)^2}{d}} \tag{2}$$

where $\hat{y}_i$ = calculated progress from an S-curve formula for x_i; y_i = actual progress at x_i; d = number of time points (progress measurements) for a project.

Chao and Chien (2009) fitted equation (1) to the actual progress data of 101 projects for the second freeway of Taiwan and that of 27 projects in Skitmore (1992) and evaluated its fitting accuracy. An average *RMSE* of less than 0.025 was achieved, which is at least as good as the widely cited Logit transformation formula in Kenley and Wilson (1986). Chao and Chien (2009) also used four project attributes, i.e. contract amount, duration, type of work, and location, as input factors in developing a neural network model and a multiple regression model for estimating the values of the polynomial's two parameters *a*, *b* for producing S-curves, which, however, achieved progress prediction accuracy at an average *RMSE* of around 0.055 and 0.058, respectively, not significantly lower than the 0.061 average *RMSE* for the average curve method using none of the above project attributes. Therefore, there is room for improvement in the model for better performance.

3. PROPOSED CHANGES IN S-CURVE MODEL

In order to seek reduction of errors in progress prediction, the previous model should be examined and changes useful for improving its accuracy are proposed, which comprise three parts: model outputs, model inputs, and method for building the model, as discussed below.

3.1. Model outputs

First, model outputs are reviewed. The previous model's outputs are the parameters of equation (1), the values of *a*, *b*. However, for an S-curve produced from given *a*, *b*, the position of the inflection point, *p*, and the slope at it, *s*, can be derived from *a*, *b*, and vice versa, as shown next.

The first derivative of equation (1) with respect to *x* is the slope of the S-curve, meaning the percent progress for a percent project time, which is:

$$y' = 3ax^2 + 2bx + (1 - a - b) \tag{3}$$

Equation (3) is the pace curve and proper values of *a*, *b* can make it convex for *x*=0~1, consistent with common projects progressing slow at the beginning and near the end but fast in the middle. The value of *x* that maximizes equation (3) is the position of the S-curve's inflection point. To solve *p*, take differentiation of equation (3) with respect to *x* and set the derivative to zero as:

$$y'' = 6ax + 2b = 0 \tag{4}$$

From equation (4), the position of the inflection point of equation (1) can be solved as:

$$p = \frac{-b}{3a} \tag{5}$$

From equations (3) and (5), the slope at the inflection point can be obtained as:

$$s = \frac{-b^2}{3a} + (1 - a - b) \tag{6}$$

In a reverse manner, *a*, *b* can be derived from equations (5) and (6) and expressed in *p*, *s* as:

$$a = -[\frac{(s-1)}{(p+\frac{1}{3p}-1)}] \div (3p) \tag{7}$$

$$b = \frac{(s-1)}{(p+\frac{1}{3p}-1)} \tag{8}$$

Therefore, *p, s* can be translated into *a, b* for producing an S-curve using equation (1). In order to use them properly, we need to establish the range of their values suitable for project progress as the basis of model development. There are two basic requirements for *p, s*: (i) $0 \leqq p \leqq 1$ due to $0 \leqq x \leqq 1$; (ii) $s>1$, because if $s=1$, equation (1) becomes a straight line $y=x$ with $a=b=0$ and if $s<1$, equation (1) gives a curve that is the opposite of the S shape for $0 \leqq x \leqq 1$ where the slope at the inflection point is the least. Moreover, if *s* has a value that is too large, an unreasonable S-curve results, in which $y<0$ or $y>1$ occurs near the start or the end. Thus, there are upper bounds of *s*, which depend on *p* and have a maximum of about 1.85 for $p=0.5$.

Of the 101 case projects in Chao and Chien (2009), five have fitted S-curves with an incorrect shape, such as $p<0$, $p>1$, or $s<1$, and hence were deleted. Of the remaining 96 S-curves, 65 (about 2/3) have $p>0.5$; the minimum *p* is 0.207 and the maximum *p* is 0.922, while the minimum *s* is 1.082 and the maximum *s* is 1.821. Their average *p* of 0.554 and average *s* of 1.412 are translated into $a=-1.592$ and $b=2.644$. However, the averages of 96 pairs of fitted *a, b* are $a=-1.488$ and $b=2.372$, which are translated into $p=0.531$ and $s=1.376$. The average S-curves from these two averaging methods are shown in Figure 1, where the differences in progress, measured by $RMSE \fallingdotseq 0.021$, are due to nonlinearity of equations (5)(6)(7)(8) for translating *a, b* and *p, s*. Later, in evaluating model prediction accuracy, both methods are to be compared with the proposed model.

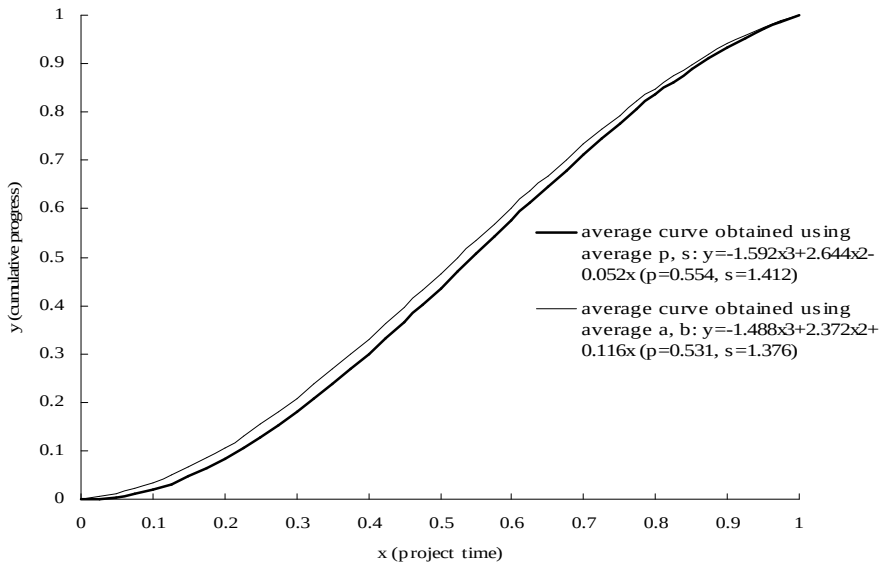

Figure 1: Average S-curves of 96 case projects in Chao and Chien (2009)

The meanings of the values of *p, s* can be seen from their relations to progress tending to be ahead of or behind schedule. Using progress at time x=0.8 for illustration, equation (1) will give various y values for different p values (0.3, 0.5, 0.7) coupled with different s values (1.2, 1.5), as shown in Table 1. Progress tends to be ahead of schedule most if p=0.3, s=1.5 and behind schedule most if p=0.7, s=1.5, while the degree of being ahead of or behind schedule is moderate if s=1.2 or p=0.5. Due to the fact that the inflection point is where project progress peaks and its slope indicates the extent of work concentration, differences in y above can be explained by the pace curve being skewed to the left or to the right, and spreading more evenly or more steeply. Hence, presumably the values of *p, s* are connected with project schedule performance metrics, e.g. schedule growth and overall average speed, and factors influencing schedule performance also influence the values of *p, s*. The present research plans to develop an improved model by incorporating relevant factors and change the model outputs from *a, b* in equation (1) to *p, s* that indicate an S-curve's key geometric feature and implies schedule performance, so as to produce an S-curve estimate and progress prediction with better accuracy.

Table 1: Values of *y* at *x*=0.8 from equation (1) for different pairs of *p, s*

p	0.3	0.3	0.5	0.5	0.7	0.7
s	1.2	1.5	1.2	1.5	1.2	1.5
y at x=0.8	0.88	0.99	0.84	0.90	0.77	0.73

3.2. Model inputs

Next, model inputs are reviewed. Chao and Chien (2009) used four project attributes, i.e. contract amount, duration, type of work, and location, as model inputs and in sensitivity analyses of the built neural network model it was found that in response to changes in inputs the S-curves generated from estimated *a, b* varied sensibly. Since the four are essential to a construction project, they are to be retained in the new model. However, using data for the previously mentioned 96 case projects, regression analyses on p and s resulted in low R^2 for both (<0.2), hence it is reckoned that some important explanatory variables are lacking.

As Konchar and Sanvido (1998) and Chao and Hsiao (2012) show, variables relating to project conditions have statistically significant influences on schedule performance and such variables include team communication, project complexity, contractor's experience, contractor's competence, consultant's competence, familiarity with work, and social environment. Since those influencing schedule performance are also likely relevant to estimation of the values of *p, s*, it is proposed that in addition to the four project attributes above, two generic factors concerning project conditions, i.e. degree of project difficulty and competence of project participants, are also used as model inputs, which are each defined by a set of variables. Influences of the two can be seen from scenarios such as delay in progress and greater concentration of work later in a project (larger *p, s*) due to higher job complexity or inexperienced contractor requiring more time in smoothing work flows, which may also be caused by shortage of resources or contractor's under-performance in procurement.

3.3. Method for building model

Lastly, the present research plans to use fuzzy inference systems (FIS) instead of neural networks (NN) for establishing the input-output relationships, because FIS has a transparent structure and how it obtains outputs from inputs is explainable unlike the opaque structure and behavior of NN. It is proposed to automatically generate more concise FIS from collected case data using the fuzzy clustering method of Chiu (1994), which has been applied to project performance prediction by (Chao and Hsiao, 2012). Next, the hybrid training method of Jang (1993) will be used to optimize the parameters of FIS. The trained FIS will then give estimates of *p, s* for producing S-curves.

4. RESEARCH WORK AHEAD

For developing a fuzzy model for S-curve estimation, data on the actual progress, project attributes, and project conditions for recently completed projects in Taiwan will be collected. For each project, equation (1) will be fitted to the standardized progress data and the pair of *a, b* solved will then be translated into *p, s* by equations (5), (6), to be used as the target outputs in training and testing FIS. The above-mentioned four attributes and two conditions factors will be used as inputs. On the latter, degree of project difficulty will be defined by complexity, familiarity, social environments etc, and competence of participants defined by team experience, technical and managerial competence, resources sufficiency, team communication etc. Since these are qualitative variables, surveys of project participants using questionnaires will be conducted for collecting their assessments. The median or average of the responses on the 1~5 scale will be taken as the value of a variable, while the arithmetic or geometric mean of the variables concerned taken as the value of a factor.

Most of the case projects collected will be used as training data and the rest reserved for testing FIS trained. Since such FIS is of single output and there are two model outputs, two FIS with the same inputs will be built for *p* and *s*. System development will use MATLAB's Fuzzy Logic Toolbox and select the options of subtractive clustering for FIS generation and the hybrid training method. For each project, the estimated *p, s* from the FIS will be translated into *a, b* by equations (7), (8), which then are used to produce an S-curve using equation (1). The accuracy of the new model in progress prediction will be evaluated based on the average training and testing *RMSE* and compared with that of the previous model, the regression model, and the average curve method.

5. CONCLUSIONS

Project progress prediction in construction involves complex cause-effect relations. The work so far shows that based on a cubic polynomial function the position of the inflection point of an S-curve and its slope are likely connected with project schedule performance and hence are proposed for use as the outputs an improved model for estimating S-curves. It is expected that the model in the form of fuzzy inference systems built using the fuzzy clustering method along with hybrid training can achieve better accuracy in progress prediction because of the revised outputs and the added inputs of project conditions factors as well as robustness of fuzzy systems. As an empirical practical tool,

the model developed from real project data would be able to give realistic progress predictions that are helpful for setting reasonable targets for effective project control.

6. ACKNOWLEDGMENTS

Support for the work presented herein from National Science Council of Taiwan is acknowledged.

REFERENCES

Blyth K and Kaka AP (2006). A novel multiple linear regression model for forecasting S-curves. Engineering, Construction and Architectural Management, 13(1), 82–95.

Chao LC and Chien CF (2009). Estimating project S-curves using polynomial function and neural networks. Journal of Construction Engineering and Management, 135(3), 169-177.

Chao LC and Hsiao CS (2012). Fuzzy model for predicting project performance based on procurement experiences. Automation in Construction, vol.28, p.71-p.81.

Chiu SL (1994). Fuzzy model identification based on cluster estimation. Journal of Intelligent and Fuzzy Systems, 2, 267-278.

Jang JSR (1993). ANFIS: adaptive-network-based fuzzy inference system. IEEE Transactions on Systems, Man, and Cybernetics, 23(3), 665-685.

Kaka AP (1999). The development of a benchmark model that uses historical data for monitoring the progress of current construction projects. Engineering, Construction and Architectural Management, Vol.6 No.3, p256-266.

Kenley R and Wilson OD (1986). A construction project cash flow model - an idiographic approach. Construction Management and Economics, Vol.4 No.3, p213-232.

Konchar M and Sanvido V (1998). Comparison of US project delivery systems. Journal of Construction Engineering and Management, ASCE, 124(6), 435–444.

Skitmore M (1992). Parameter prediction for cash flow forecasting models. Construction Management and Economics,Vol.10, p397-413.