REVIEW

CrossMark

The association analysis for risk evaluation of significant delay occurrence in the completion date of construction project

H. Anysz[1] · B. Buczkowski[1]

Abstract

It is widely known that a construction project can be described as a success if it is completed on time and without cost overrun. As the strong correlation between delay in completion date of construction projects and cost overruns exists, it is necessary to manage the risk of a delay from the very beginning: even before the start of works execution. The data concerning 139 construction contracts of the highway sections already completed, as well as their completion dates, were collected. The twelve types of data were chosen based on the criterion of time of appearance: they should be known before the start of construction works. Most of the contracts' properties have continuous values, and these had to be converted to the binary values. With the use of the association analysis, it is possible to find the combination of properties of the construction contracts that have produced significant delays in completion date in large part of analysed and already completed contracts. For this purpose, appropriate software has been created. This paper will comprise the software approach to the problem as well as the result of the analysis, as it is based on real data completed. Getting the information on which combination of construction contract properties can cause a significant delay in completion date of the contract, the client can make decisions leading to avoiding risk, i.e. providing better circumstances for project success.

Introduction

There are two critical, interrelated issues determining the success of a construction project (Skorupka et al. 2012): keeping the completion date as scheduled and the lack of the cost overrun. The delayed completion date of construction contract makes contractor cost much higher than expected (Anysz and Książek 2012). The negative impact of this delay—financial and non-financial—is not limited to the contractor and the client. The community—that the planned object will serve for—suffers from the delay too (Anysz 2017). That is why identifying the most critical causes of delays and avoiding them is so important. Different methods are applied for identifying the factors causing delays and for the validation of their importance (Głuszak and Leśniak

2015; Ibadov 2016). Both parts of the information are needed to avoid unnecessary delays. Lowering the possibility of delay occurrence can concentrate on proper planning of work execution [planning the duration of works execution (Juszczyk 2014), scheduling (Krzemiński 2016)] or on avoiding unfavourable circumstances for project execution.

The client can estimate the duration of the construction project with the use of the critical path method (popularized by project management body of knowledge, PMBoK 2000)—a deterministic method of scheduling. Probabilistic methods of scheduling (e.g. PERT—program evaluation and review technique) are not so widely applied. Introduced by Goldratt (1997) the time buffers in his critical chain method have improved preciseness of a construction project duration planning. The other popular tool for predicting the project duration is earned value method (Webb 2003). This method is obligatory in public procurements in many countries, and it is still developed, e.g. (Vandervoorde and Vanhoucke 2006). The big disadvantage of EV method is the necessity of advancing the project to perform the duration estimation. The problem of construction project scheduling is so important that Zadeh's theory (Zadeh 1999) is being implemented

✉ H. Anysz
  h.anysz@il.pw.edu.pl

[1] Civil Engineering Department, Warsaw University
  of Technology, Warsaw, Poland

Springer

too, e.g. (Ibadov 2016, 2018). The other approach (except scheduling) to predicting the duration of the construction site is to analyse factors influencing the total time of a project. Variety of tools are applied, e.g. a case-based reasoning (Koo et al. 2010), artificial neural networks (Jha and Chockalingam 2011; Anysz 2017), classification and regression trees (CART) models, multiple regression (Czarnigowska and Sobotka 2013). In the aforementioned models, selected factors influencing the total time of the project (independent variables) are the base to estimate the duration of the project (independent variable).

As the contractors base their decisions on their experience (PMBoK 2000), carrying out decisions about participation in a given tender procedure (Leśniak 2015), the client can analyse already completed projects and try to avoid circumstances that resulted in a significant delay in completion of the projects in the past. The novelty of this paper is in analysing the coincidences of phenomena, which describe analysed construction projects (by factors' or parameters' values) to find their unfavourable combination, i.e. the combination that very often was accompanying a significant delay of project completion. A very suitable tool to find the set of unfavourable circumstances is association analysis (Morzy 2013).

Completing the database

The subject of the research

The highways and express roads are homogeneous-type building objects. They are being built based on public procurement law introduced in Poland. There is one client for all of them: the General Directorate for National Roads and Highways GDDKiA. All projects of building express roads and highways completed in Poland between January 2009 and December 2013 have been analysed.

The choice of parameters describing the projects

To find the reasons for delays in completion dates of construction projects, own research has been done (Anysz 2013) as well as own analysis (Anysz 2012).

The review of international literature focused on the reasons of delays of the construction projects has shown that keeping the time schedule is the worldwide problem (Chan and Kumaraswamy 1997; Doloi et al. 2012; Faridi and El-Sayegh 2006; Frimpong et al. 2003; Kazaz et al. 2012; Niazai and Gidado 2012; Ogunlana et al. 1996; Sambasivan and Soon 2008; Sweis et al. 1998; Toufic et al. 1998). The result of the research mentioned above was the list of 142 possible reasons for delays (Anysz 2017). A

considerable number of them was reduced, mainly according to the fact that the moment of analysis is before the choice of a contractor (by the client), before the start of building works.

Four groups of reasons causing delays in completion dates of construction projects were excluded from further analysis:

- the reasons that arise during the works execution (e.g. machine failure),
- the reasons arising before the start of building works that can be detected during the works execution (e.g. inconsistency in design documentation),
- the reasons that do not differentiate the projects (e.g. bureaucracy within the client's organization; it was the same client for all analysed projects),
- the reasons not currently existing in Poland (e.g. military threats).

These several possible causes of delays for further analysis are included in Table 1. Label D was left for marking the delay. Its integer value was calculated for each project based on the formula:

$$D_i = \begin{cases} T_i^{(\mathrm{r})} \le T_i^{(\mathrm{pl})} \to 0 \\ T_i^{(\mathrm{r})} > T_i^{(\mathrm{pl})} \to T_i^{(\mathrm{r})} - T_i^{(\mathrm{pl})} \end{cases} \tag{1}$$

where $T_i^{(\mathrm{pl})}$ planned duration of the project given in days, $T_i^{(\mathrm{r})}$ observed real duration of the project given in days, i number of analysed project.

The sources of information

The twelve factors mentioned above that may influence the delay of the completion date of road construction projects can be categorized into three main groups by origin:

- the client decision dependent (B, C, E, H),
- the contractor dependent (A, G, L, M),
- macroeconomic factors (I, J, K).

The one not included (F) arises from the technical matters and the standing of the national economy. The majority of data has been provided by GDDKiA on request of Warsaw University of Technology. Macroeconomic factors have been found in Polish Central Statistical Office (GUS). For the real completion dates approx. 500 websites have been checked. The data concerning the number of employees and the yearly sales of contractors were obtained commercially. The complete set of twelve feature values was

Table 1 Possible causes of delays and their values

Label	The cause of delay	Values
A	Value of works	Rational number (PLN)
B	Length of the section built	Rational number (km)
C	Planned duration of the project	Integer number (days)
E	Project scope	Binary: design and build = 1; build = 0
F	Project type	Binary: build = 1; modernize, etc. = 0
G	The total, average number of employees employed by contractor**	Integer number (person)
H	Half of the year of works commencement	Binary: first half = 0; second half = 1
I	The trend of the unemployment rate in Poland*	Binary: decreasing = 0; increasing = 1
J	The trend of price index in Polish construction industry*	Binary: decreasing = 0; increasing = 1
K	The trend of total sales in Polish construction industry*	Binary: decreasing = 0; increasing = 1
L	Number of partners in consortium (acting as contractor)	Integer number
M	Summarized yearly total sales of consortium partners**	Rational number (PLN)

*Calculated year to year (the year preceding the commencement of works, to the year before)

**Calculated for the year preceding the commencement of works

gathered for 139 projects, and only these were analysed further.

The aim of analysis

Definitions

Let X and Y are the features characterizing the process that was repeated multiple N times. Let X be the predecessor (often called a *body*), and Y be consequent (often called a *head*) (Morzy 2013). The level of support of Y by a body (X) can be defined as:

$$\sup(X \rightarrow Y) = \frac{n(X \cap Y)}{N} \tag{2}$$

where $n(X \cap Y)$ number of cases (processes) where X and Y are present simultaneously, N the total number of cases (processes) analysed.

The confidence of the rule "if X then Y" can be calculated as:

$$\operatorname{conf}(X \rightarrow Y) = \frac{n(X \cap Y)}{n(X)} \tag{3}$$

where $n(X)$ number of cases (processes) from N total number of processes, where appears (Lasek and Pęczkowski 2013).

The *body* (X) can also be a set of *bodies*. Then support and the confidence are calculated for the set of *bodies* where all of them appear simultaneously $(X_1 \cap X_2 \cap X_3 \cap \cdots \cap X_m) \rightarrow Y$.

The aim

This paper aims to find which combination of causes of delays (listed in Table 1, describing completed projects) gives the highest confidence of significant delay in completion date of the project. Secondly, the strength of the rule found (Larose and Larose 2016) is in focus. The client, getting the information, e.g. that 12% of the contract they have executed in the last 4 years had the following parameters:

- planned duration of the project was below 300 days,
- it was more than 2 partners in contractor's consortium,
- the length of the section of road planned to built was longer than 15 km

(i.e. sup = 0.12) and that on each project (fulfilling the three conditions above) delay in completion date was longer than 130 days, can change initial parameters of the next project of this type hoping to avoid a significant delay. Having such a strong warning, the client can search for other initial parameters and check *sup* and *conf* for a new combination of parameters. If the confidence of a significant delay appearance were much lower, it could be deduced that the risk of a significant delay appearance in a planned project would be lowered.

The software for association analysis

The proposed type of analysis is for binary bodies. Six of them (from Table 1) are binary already (E, F, H, I, J, K). The rest, which has integer or rational values, has to be digitized. For each cause of delay, a threshold should be set. For instance, for the value of contract (A), setting the threshold a will divide the set of contract values into two subsets in the following way:

$$a_i = \begin{cases} A_i \leq a \rightarrow 0 \\ A_i > a \rightarrow 1 \end{cases} \tag{4}$$

where A_i the value of i contract before digitization, and a_i the value of i contract after digitization.

The set of delay (*D*) values should be digitized as well. The contribution of the second author was to create a software (applying algorithm provided by the first author) allowing for the digitization of input data set, and based on that, analysing all possible combinations of bodies (from 1 to 12 bodies). Except for the input data set, the seven thresholds must be given as an input—for initially non-binary data types and delays. As an output 12 tables are provided by the software containing all combination of causes of delays with *sup* and *conf* calculated (each table for the separate number of causes of delays). The *R* environment was utilized (Chambers 2008; Cotton 2013) (R x64 3.4.1) with *RStudio version 1.0.153*. The authors have named the software *ABAAC v.1.0* (*Anysz-Buczkowski-Association-Analysis-in-Construction*).

The thresholds for bodies were set to median values. It was assumed that delay in completion date of road construction projects longer than 100 days is significant, so the threshold for the head (*D*) was $d = 100$. It has to be stated that the ABAAC software analyses each body for both its values: 0 or 1. When, for instance, a subset of low contract value influences the delay more than the high contract value, ABAAC will mark this feature with lower index r (as reverse).

Results of the rules finding

The confidence and support were calculated for all combinations of bodies. It has occurred that 1384 of them had confidence equal to 1, but only three of them had maximum support equal to 8,6%. They are shown in Table 2.

Let us analyse the results (from Table 2) with the lowest number of bodies (C_r–*E*–*J*–*L*). The achieved confidence equal to 1 and support equal to 8.6% mean that when:

- the planned time of work execution was not longer than 670 days and
- the contractor has built the section of a road basing on the design provided by the client and
- before the project has started, the trend of price index in the construction industry was increasing and
- the contractor's consortium consisted of more than two companies,

then (in every case when the combination of bodies C_r–*E*–*J*–*L* has appeared) a delay of the completion date of the project

Table 2 Combinations of bodies with conf = 1 and the highest support

Combinations of bodies	Support (%)	Confidence (%)
C_r–*E*–*J*–*L*	8.6	100
C_r–*E*–*F*–G_r–*J*	8.6	100
A–*E*–*F*–I_r–*J*–*K*	8.6	100

was always over 100 days. It is a strong warning against longer than planned works execution. The client cannot change the price index in the construction industry, but the other bodies can be changed. The effect of that is shown in Fig. 1.

The least risky case is when the planned time is assumed longer, and the contractor's consortium has less than 3 partners. There were 26.6% (sup $((E \cap J) \to D) = 0.266$) of such cases, and a significant delay has occurred in 66.1% of them (conf($(E \cap J) \to D) = 0.661$). There are more possible decisions where confidence is approx. 0.75 and support is between 8.6 and 26.6% as can be seen in Fig. 1.

The combinations of bodies with the highest support and conf > 0.75 have also been found. (There were only three such cases.) The possible client's decisions are shown in Fig. 2. They were calculated for the combination of bodies having the highest support (*A*–*E*–*K*) among combinations with confidence higher than 75%. Extending the database beyond 2013 and keeping it updated would serve the public client as GDDKiA to assess—with the use of proposed method—what is the probability of not keeping the time planned for a given road construction project. It is not recommended to apply the association analysis to estimate the duration of the project solely (Ahmed et al. 2015) but rather as an auxiliary tool assessing the risk level of the chosen solution. The detection of the high risk (the high confidence of the occurrence of the significant delay) should be accompanied by a proposal of the reaction to the high risk detected (Ibadov 2017). The proposed tool suggests the solution by showing the combinations of unfavourable factors causing delays, where the confidence of the significant delay occurrence is much lower.

Conclusion

The original software ABAAC—created for this paper purposes—has allowed for the analysis of real data concerning the combinations of causes of delays in road
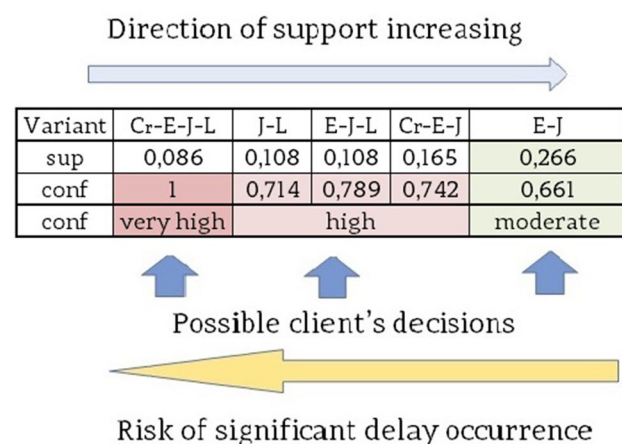

Fig. 1 Possible client's decisions where the risk of a significant delay is evaluated

Direction of support increasing

Direction of confidence increasing

Variant	A-E-K	A-K	E-K	K
sup	0,223	0,245	0,374	0,403
conf	0,756	0,667	0,612	0,557

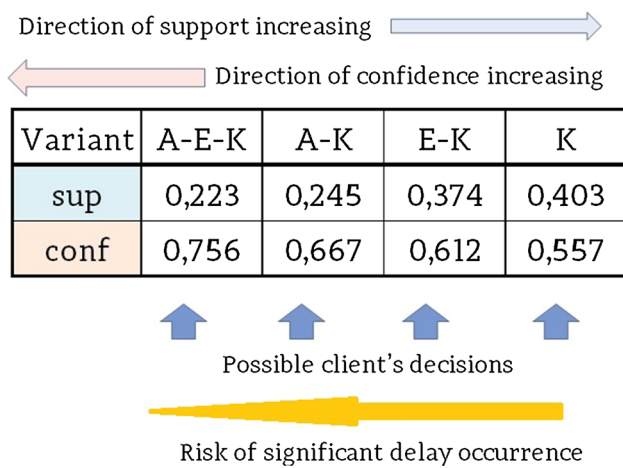Possible client's decisions

Risk of significant delay occurrence

Fig. 2 Possible client's decisions where the risk of a significant delay is evaluated

construction projects in Poland. This innovative use of association analysis enabled the rules to be discovered: Which phenomena (the causes of delays) appearing simultaneously before the commencement of works (even before a contractor is chosen) can make the duration of the project much longer than planned. Some of the rules found show 100% confidence.

These kinds of warnings cannot be treated as predictions itself. They can be an auxiliary tool verifying the assumed duration of the project (evaluated with the use of the traditional methods). According to the database completed, the proposed tool can be used for express roads and highways being built in Poland. The results of association rules found, having the confidence of 100% (with rather low support) or the rules with still high confidence (75%) and moderate support (above 20%), prove that application of association analysis in construction industry helps to discover unfavourable initial conditions or decisions that have accompanied significant delays in projects already completed. This tool can support the risk evaluation and helps to set the reaction to the risk detected. The strong association rules found can be directly used in the construction industry as they were based on real data. However, even more important finding is that combinations of *bodies* produce much higher confidence with a significant delay than the confidence of a single *body*. This constitutes a proof of successful application of association analysis in organizational matters of the construction industry, so the tool and software will be developed. The application of the proposed type of analysis for other branches of construction industry or in other states is possible but requires completing historical data about the project already completed.

References

A Guide to the Project Management Body of Knowledge (PMBOK® Guide) (2000) Project Management Institute, Pennsylvania USA, 2000 Edition

Ahmed AM, Bakara AA, Hamdana AR, Syed Abdullah SM, Jaafarb O (2015) Sequential pattern discovery algorithm for Malaysia rainfall prediction. Acta Physica Pol A. https://doi.org/10.12693/APhysPolA.128.B-324

Anysz H (2012) Analiza zależności opóźnień w realizacji kontraktów na wykonanie robót budowlanych od czynników leżących poza sferą organizacji budowy i robót budowlanych, Teoretyczne podstawy Budownictwa 69–82. Oficyna Wydawnicza PW, Warsaw

Anysz H (2013) Przyczyny powstawania opóźnień w realizacji kontraktów budowlanych – analiza wstępnych wyników badania ankietowego, AUTOBUSY, Technika, Eksploatacja, Systemy Transportowe 963-972 3/2013 ISSN 1509-5878

Anysz H (2017) Delays in construction contracts execution. The evaluation of the possibility of their occurrence with the use of artificial neural networks. Ph.D. Thesis, Oficyna Wydawnicza PW, Warsaw, Poland. https://doi.org/10.13140/rg.2.2.14129.68960

Anysz H, Książek M (2012) Wpływ opóźnień w realizacji kontraktu budowlanego na wzrost kosztów wykonawcy, Teoretyczne podstawy Budownictwa 83–96. Oficyna Wydawnicza PW, Warsaw

Chambers JM (2008) Software for data analysis. Programming with R. Springer, Berlin. https://doi.org/10.1007/978-0-387-75936-4

Chan DWM, Kumaraswamy MM (1997) CA comparative study of causes of time overruns in Hong Kong construction projects. Int J Proj Manag 15:55–63

Cotton R (2013) Learning R. A step-by-step function guide to data analysis. O'Reilly, Newton. ISBN 978-1-449-35710-8

Czarnigowska A, Sobotka A (2013) Zależność czas–koszt w przewidywaniu czasu realizacji budowy. Budownictwo i Architektura 12:23–30

Doloi H, Sawhney A, Iyer KC, Rentala S (2012) Analysing factors affecting delays in Indian construction projects. Int J Proj Manag 30:479–489

Faridi AS, El-Sayegh SM (2006) Significant factors causing delay in UAE construction industry. Constr Manag Econ 24:1167–1176

Frimpong Y, Oluwoye J, Crawford L (2003) Causes of delay and cost overruns in construction of groundwater projects in a developing countries. Ghana as a case study. Int J Proj Manag 21:321–326

Głuszak M, Leśniak A (2015) Construction delays in clients opinion—multivariate statistical analysis. Procedia Eng 123:182–189. https://doi.org/10.1016/j.proeng.2015.10.075

Goldratt EM (1997) Critical chain. The North River Press, Great Barrington

Ibadov N (2016) Determination of the risk factors impact on construction projects implementation using fuzzy sets theory. Acta Physica Pol A 130:107–111. https://doi.org/10.12693/aphyspola.130.107

Ibadov N (2017) Selection of construction project taking into account technological and organizational risk. Acta Physica Pol A 132:974–977. https://doi.org/10.12693/APhysPolA.132.974

Springer

Ibadov N (2018) Construction project planning under fuzzy time constraint. In: Construction project planning under fuzzy time constraint. https://doi.org/10.1007/s13762-018-1695-x

Jha KN, Chockalingam CT (2011) Prediction of schedule performance of Indian construction projects using an artificial neural network. Constr Manag Econ 29(9):901–911. https://doi.org/10.1080/01446193.2011.608691

Juszczyk M (2014) A concise review of methods of construction works duration assessment. Technical Transactions 2014 (Budownictwo Zeszyt 2-B 2014), pp 193–202

Kazaz A, Ulubeyli S, Tuncbilekli NA (2012) Causes of delays in construction projects in Turkey. J Civ Eng Manag 18(3):426–435

Koo C, Hong T, Hyun C (2010) A CBR-based hybrid model for predicting a construction duration and cost based on project characteristics in multi-family housing projects. Can J Civ Eng. https://doi.org/10.1139/l10-007

Krzemiński M (2016) KASS v. 2.2. scheduling software for construction with optimization criteria description. Acta Physica Pol A 130:1439–1442. https://doi.org/10.12693/APhysPolA.130.1439M

Larose DT, Larose CD (2016) Discovering knowledge in data. Willey IEEE, Hoboken. ISBN 978-81-265-5834-6

Lasek M, Pęczkowski M (2013) Analiza asocjacji i reguły asocjacyjne w badaniu wyborów zajęć dydaktycznych dokonywanych przez studentów. Zastosowanie algorytmu Apriori, Ekonomia. Rynek. Gospodarka. Społeczeństwo nr 34/2013, Warsaw, Poland

Leśniak A (2015) Classification of the bid/no bid criteria—factor analysis. Arch Civ Eng LXI(4):79–90. https://doi.org/10.1515/ace-2015-0037

Morzy T (2013) Eksploracja danych. Metody i algorytmy, Wydawnictwo Naukowe PWN, Warsaw, Poland

Niazai GA, Gidado K (2012) Causes of project delay in the construction industry in Afghanistan. In: EPPM2012, 10–11th September, University of Brighton, Brighton

Ogunlana SO, Promkuntong K, Jearkjirm V (1996) Construction delays in a fast-growing economy: comparing Thailand with other economies. Int J Proj Manag 14(1):35–45

Sambasivan M, Soon YW (2008) Causes and effects of delays in Malaysian construction industry. Int J Proj Manag 25:517–526

Skorupka D, Kuchta D, Górski M (2012) Zarządzanie ryzykiem w projekcie. WSOWL, Wrocław

Sweis G, Swies R, Abu Hammad A, Shboul A (1998) Delays in construction project: case of Jordan. Int J Proj Manag 28:665–674

Toufic M, Tawil M, Tawil W (1998) Causes of delays in the construction industry in Lebanon. Eng Constr Archit Manag 5:252–260

Vandervoorde S, Vanhoucke M (2006) A comparison of different project duration forecasting methods using earned value metrics. Int J Proj Manag 24:289–302

Webb A (2003) Using earned value: a project manager's guide. Gower Publishing, Farnham

Zadeh LA (1999) Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets Syst 100(1):9–34. https://doi.org/10.1016/S0165-0114(99),80004-9