6R-08

Toward a Delay Risk Analysis and Prediction System on Road Construction Projects in Zimbabwe

Florence GUNDIDZA † Helton YAWOVI† Masato KIKUCHI† Tadachika OZONO†

NAGOYA INSTITUTE OF TECHNOLOGY†

1. INTRODUCTION

1.1. Research Background

The construction industry plays an important role in establishing the infrastructure required for socio-economic development, besides directly contributing to economic growth. However, construction project delay is a global phenomenon. Project delay can be defined as an action or event that extends the time required to complete the project identified in a contract [1]. Project delay has an adverse effect on project performance, which leads to cost overruns and productivity reduction. Its occurrence is mainly attributed to the interdependent imminent risk factors and uncertainties associated with the complex and dynamic nature of construction processes. Such factors may include stakeholder(s) incompetence, poor communication, inadequate estimation of employed resources, contractual deviations, or even municipal constraints [1].

To provide accurate estimates of project durations, construction firms typically adopt standard quantitative delay risk analysis tools which are used to investigate the complete extent of risk associated with scheduled work items to estimate more reasonable project completion dates. These tools are data-intensive and require estimates which are obtained from historical data of similar projects or in cases of absence of such data expert opinion is used. The latter is subjective and poses several key limitations. For example (1) using imprecise and/or ambiguous data sources that would typically add another layer of uncertainty to the analysis; (2) overlooking the complex and interdependent nature of inherent risk factors in construction projects that might significantly influence original predictions, and (3) using data that are rarely updated to represent the actual project progressions since the subjective data collected are relevant only at the time of apprehension.

In that regard, the potential of machine learning (ML) techniques and algorithms in analyzing voluminous, complex, and interdependent datasets of varying structures for deriving useful insights cannot be overemphasized.

Toward a Delay Risk Analysis and Prediction System on Road Construction Projects in Zimbabwe

†Florence GUNDIDZA †Helton YAWOVI ‡Masato KIKUCHI ‡Tadachika OZONO

‡Dept of Computer Science, Graduate School of Engineering, Nagoya Institute of Technology

1.2. Literature Review

Recently, artificial intelligence (AI) models have been widely applied in different fields of engineering and science. Several studies used AI models in their research on construction project management but only a few emphasize risk prediction and classification. Asadi et al. (2015) used a decision tree and a Naive Bayes model based on a questionnaire survey to predict delays in construction logistics [2]. Naji et al. (2018) used a Bayesian decision tree model to predict the impact of contract changes on the time and quality performance of construction projects [3]. Gondia et al. (2020) utilized Naive Bayes and decision tree models to predict the delay risk in construction projects [4].

A literature survey that we conducted showed that a few studies have focused on applications of AI techniques within Zimbabwe's Road Construction research in general and, to an even much lesser extent, on delay risk analysis.

1.3. Research Objectives

To achieve the ultimate goal which is to develop an efficient predictive tool to analyze and predict delay risk problems based on objective delay sources established from previous construction projects data, in this paper, we subsequently aim to build a relevant construction project delay risk dataset as no similar data set is available in open-source literature. The main contribution of the investigation is to establish an accurate dataset that can be adopted in the development of an effective tool for the prediction of future durations and monitoring risk levels, based on these projects.

2. METHODOLOGY

In a bid to achieve the aim of building an accurate data set, subsequently, relevant sources and factors of delay risk are extracted from the literature and adopted by the industry. A questionnaire survey is then adopted to measure the impact of various sources on the delay level in construction projects. Due to unforeseen circumstances i.e. COVID19 restrictions, we deem this as a limitation to conduct a physical survey, rather Natural Language Processing (NLP) model "Bidirectional Encoder Representations from Transformers (BERT)" is adopted to carry out mainly two tasks [5].

Task 1: Detection of Delay Risk

Initially, a BERT language model with a binary classification head is trained to predict whether a sentence was written in project reports regarding project delay or not. The data collected and used for training and testing the model are project reports obtained from reliable sources like the African Development Bank, Asian Development Bank, and JICA. A maximum length of tokens per sentence, a training rate, and a validation size are established. Parameters to determine if the system should trigger an alert given a list of known project report texts are defined: the minimum average probability threshold θ, the minimum number of texts necessary to trigger an alert, and the maximum number of texts that the system will take into account to make its decision. Given a growing list of texts from the reports, the system will trigger an alert if the average probability of the known texts for that project is greater or equal than θ, the number of known texts is greater or equal to the minimum, and lower or equal to the maximum. The classifier is then tested using 30% of the dataset.

Task 2: Measuring the Impact of delay sources on the delay level

This task consists of automatically estimating the level of impact of several delay sources on Road Construction Projects. For that, a questionnaire with questions related to different delay sources is provided. Each question has five possible answers which are related to different levels of severity as indicated in Table 1. Both training and test datasets are positive reports from task 1. For this task, a similar method as the one employed in Task 1 is used, but the problem is treated as a multi-class labeling problem. The base language model with a head for multi-class classification for every question is fine-tuned. The classifier is trained for an optimal number of epochs. The maximum sentence length and the learning rate to train the model are then established. 20% of the training data is assigned for validation. To predict the questionnaire answers for a given question and the associated classifier, a softmax prediction vector is used for every text written and sum them up. The class with the highest accumulated value is the answer to the questionnaire predicted.

To measure the performance of the system for task 1, standard classification measures precision (P), recall (R), and F1 are computed with respect to the positive class. For task 2 Average Hit Rate (AHR) here HR is the ratio of cases where the automatic questionnaire has the same answer as the actual questionnaire. Delay Source Category Hit Rate (DSCHR) computes the fraction of cases where the automated questionnaire led to a Delay Source Impact category (out of 5 categories: very little, little, moderate, severe, and very severe) that is equivalent to the depression category obtained from the real questionnaire.

Table 1: A factoid question with five candidate answers.

Q	**How much change does delay of progress payment on the project output**
A_1	There was very little change in the actual outputs, these overruns in project period cannot be judged to be appropriate in view of the actual outputs.
A_2	Due to delays in payment by the project owner, there was a little change in the project performance.
A_3	The project performance was moderately longer than planned.
A_4	Delays in the payment of the government share were significant and did affect the contractors' cash flow and project progress.
A_5	There was very significant change in project performance due to delays release of funds

3. CONCLUSION

In this paper we describe how BERT is employed for tasks 1: delay risk detection on Road construction projects using past projects reports; and Task 2: inferring answers to a delay source-based questionnaire survey. The results of task 2 are anticipated to be used in the ultimate goal of developing an efficient predictive tool to analyze and predict delay risk problems based on objective delay sources.

ACKNOWLEDGEMENT

This work was supported in part by JSPS KAKENHI Grant Number JP19K12266.

REFERENCES

1. Aziz, R. F. **2013**. "Ranking of delay factors in construction projects after Egyptian revolution." Alexandria Eng. J. 52 (3): 387–406.
2. Asadi, A.; Alsubaey, M.; Makatsoris, C. **2015**. "A machine learning approach for predicting delays in construction logistics." Int. J. Adv. Logist. 4, 115–130.
3. Naji, H.; Ibrahim, A.M.; Hassan, Z. **2018** "Evaluation of Legislation Adequacy in Managing Time and Quality Performance in Iraqi Construction Projects- a Bayesian Decision Tree Approach." Civ. Eng. J. 4, 993.
4. Gondia, A.; Siam, A.; El-Dakhakhni, W.; Nassar, A.H.**2020**. "Machine Learning Algorithms for Construction Projects Delay Risk Prediction." *J. Constr. Eng. Manag. 146*, 04019085.
5. Mart´ınez-Casta˜no,R.; Htait, A.; Azzopardi,L.; Moshfeghi, Y. **2020**. "Early risk detection of self-harm and depression severity using bert-based vtransformers: ilab at clef erisk 2020," Early Risk Prediction on the Internet 2020.