UNIVERSITY OF CAMBRIDGE
Department of Engineering

CONSTRUCTION ENGINEERING MASTERS DISSERTATION ABSTRACT

Improving Construction Schedule Risk Analysis with Machine Learning

The construction industry has been forever blighted by delay and disruption. More than 50% of all construction projects undertaken in the UK finish later than planned, causing damage to the interests of stakeholders. Given the high level of complexity and non-uniformity associated with large infrastructure projects, they are particularly susceptible to delay. Planning for the uncertainty inherent in these projects is a particularly challenging practice, the successful completion of which relies on several important factors including the management of complexity and risk, competing contract priorities, optimism bias and strategic misrepresentation, moral hazard, and skills shortages. In a highly subjective environment such as competitive contract bidding, it can be difficult to provide a realistic baseline plan which manages all of the competing delay hazards. This research seeks to provide a framework for improving the accuracy of baseline schedules, using the data created during the commission of c.£100bn worth of infrastructure construction work carried out in the UK each year.

There are a number of existing techniques and technologies available to assist planners in undertaking a meaningful risk analysis on their construction schedules. However, these tend to depend quite heavily on the accurate estimation of risk boundaries for each task, which often makes the task of carrying out such an analysis prohibitively onerous. Although the Monte Carlo method and its variants (e.g. casebased Monte Carlo method and agent-based Monte Carlo method) are the most widely used risk analysis techniques, these existing methods are limited in their scope and by their subjectivity – they only analyse a small number of projects or are based on expert opinion alone.

Using a database of 444,173 tasks from a diverse sample of 302 completed infrastructure construction projects, this research focuses on the use of machine learning to create an objective system of infrastructure schedule risk analysis. This quantitative approach seeks to address the gaps in knowledge by leveraging the value of completed project schedule data and understanding some of the factors behind inaccurate schedule estimation. The author found that a combined approach, using Empirical Bayesian Networks, to generate uncertainty and covariance, with a supervised machine learning technique – Support Vector Machine regression – delivered promising results on a sample of ten trial projects. This method is compared with a common and comparable state-of-practice technique (Monte Carlo Simulation) and is found to offer significant improvements in simulation accuracy and representativeness.

John Fitzsimmons

June 2020