Tunnelling and Underground Space Technology 27 (2012) 86–100

Contents lists available at SciVerse ScienceDirect

Tunnelling and Underground Space Technology

journal homepage: www.elsevier.com/locate/tust

Risk analysis during tunnel construction using Bayesian Networks: Porto Metro case study

Rita L. Sousa, Herbert H. Einstein *

Dept. of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, USA

ARTICLE INFO

Article history:
Received 16 December 2010
Received in revised form 15 July 2011
Accepted 17 July 2011
Available online 27 August 2011

Keywords:
Risk
Tunneling
Bayesian Networks

ABSTRACT

This paper presents a methodology to systematically assess and manage the risks associated with tunnel construction. The methodology consists of combining a geologic prediction model that allows one to predict geology ahead of the tunnel construction, with a construction strategy decision model that allows one to choose amongst different construction strategies the one that leads to minimum risk. This model used tunnel boring machine performance data to relate to and predict geology. Both models are based on Bayesian Networks because of their ability to combine domain knowledge with data, encode dependencies among variables, and their ability to learn causal relationships. The combined geologic prediction–construction strategy decision model was applied to a case, the Porto Metro, in Portugal. The results of the geologic prediction model were in good agreement with the observed geology, and the results of the construction strategy decision support model were in good agreement with the construction methods used. Very significant is the ability of the model to predict changes in geology and consequently required changes in construction strategy. This risk assessment methodology provides a powerful tool with which planners and engineers can systematically assess and mitigate the inherent risks associated with tunnel construction.

1. Introduction

There is an intrinsic risk associated with tunnel construction because of the limited a priori knowledge of the existing subsurface conditions. Although the majority of tunnel construction projects have been completed safely there have been several incidents in various tunneling projects that have resulted in delays, cost overruns, and in a few cases more significant consequences such as injury and loss of life. It is therefore important to systematically assess and manage the risks associated with tunnel construction. A detailed database of accidents that occurred during tunnel construction was created by Sousa (2010). The database contains 204 cases all around the world with different construction methods and different types of accidents. The accident cases were obtained from the technical literature, newspapers and correspondence with experts in the tunneling domain.

Knowledge representation systems (or knowledge based systems) and decision analysis techniques were both developed to facilitate and improve the decision making process. Knowledge representation systems use various computational techniques of AI (artificial intelligence) for representation of human knowledge and inference. Decision analysis uses decision theory principles supplemented by judgment psychology (Henrion, 1991). Both emerged from research done in the 1940s regarding development of techniques for problem solving and decision making. John von Neumann and Oscar Morgensten, who introduced game theory in "Games and Economic Behavior" (1944), had a tremendous impact on research in decision theory.

Although the two fields have common roots, since then they have taken different paths. More recently there has been a resurgence of interest by many AI researchers in the application of probability theory, decision theory and analysis to several problems in AI, resulting in the development of Bayesian Networks and influence diagrams, an extension of Bayesian Networks designed to include decision variables and utilities. The 1960s saw the emergence of decision analysis with the use of subjective expected utility and Bayesian statistics. Howard Raiffa, Robert Schlaifer, and John Pratt at Harvard, and Ronald Howard at Stanford emerged as leaders in these areas. For instance Raiffa and Schlaifer's Applied Statistical Decision Theory (1961) provided a detailed mathematical treatment of decision analysis focusing primarily on Bayesian statistical models. Pratt et al. (1964) developed basic decision analysis. while Eskesen et al. (2004) and Hartford and Baecher (2004) provide good summaries on the different techniques (fault trees, decision trees, etc.) that can be used to assess and manage risk in tunneling.

* Corresponding author. Address: 70 Massachusetts Ave., Room 1-342, Cambridge MA 02139, USA. Tel.: +1 617 253 3598; fax: +1 617 253 6044.

E-mail address: einstein@mit.edu (H.H. Einstein).

0886-7798/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.tust.2011.07.003

Various commercial and research software for risk analysis during tunnel construction have been developed over the years, the most important of which is the DAT (Decision Aids for Tunneling), developed at MIT in collaboration with EPFL (Ecole Polytechnique Fédérale de Lausanne). The DAT are based on an interactive program that uses probabilistic modeling of the construction process to analyze the effects of geotechnical uncertainties and construction uncertainties on construction costs and time. (Dudt et al., 2000; Einstein, 2002) However, the majority of existing risk analysis systems, including the DAT, deal only with the effects of random ("common") geological and construction uncertainties on time and cost of construction. There are other sources of risks, not considered in these systems, which are related to specific geotechnical scenarios that can have substantial consequences on the tunnel process, even if their probability of occurrence is low.

This paper attempts to address the issue of specific geotechnical risk by first developing a methodology that allows one to identify major sources of geotechnical risks, even those with low probability, in the context of a particular project and then performing a quantitative risk analysis to identify the "optimal" construction strategies, where "optimal" refers to minimum risk. For that purpose a decision support system framework for determining the "optimal" (minimum risk) construction method for a given tunnel alignment was developed. The decision support system consists of two models: a geologic prediction model, and a construction strategy decision model. Both models are based on the Bayesian Network technique, and when combined allow one to determine the 'optimal' tunnel construction strategies. The decision model contains an updating component, by including information from the excavated tunnel sections. This system was implemented in a real tunnel project, the Porto Metro in Portugal.

2. Background on Bayesian Networks

Bayesian Networks are graphical representations of knowledge for reasoning under uncertainty. They can be used at any stage of a risk analysis, and may substitute both fault trees and event trees in logical tree analysis. While common cause or more general dependency phenomena pose significant complications in classical fault tree analysis, this is not the case with Bayesian Networks. They are in fact designed to facilitate the modeling of such dependencies. Because of what has been stated, Bayesian Networks provide a good tool for decision analysis, including prior analysis, posterior analysis and pre-posterior analysis. Furthermore, they can be extended to influence diagrams, including decision and utility nodes in order to explicitly model a decision problem.

A Bayesian Network is a concise graphical representation of the joint probability of the domain that is being represented by the random variables, consisting of (Russel and Norvig, 2003):

- A set of random variables that make up the nodes of the network.
- A set of directed links between nodes. (These links reflect cause–effect relations within the domain.)
- Each variable has a finite set of mutually exclusive states.
- The variables together with the directed links form a directed acyclic graph (DAG).
- Attached to each random variable A with parents $B_1, \ldots, B_n$ there is a conditional probability table $P(A = a | B_1 = b_1, \ldots, B_n = b_n)$, except for the variables in the root nodes. The root nodes have prior probabilities.

Fig. 1 is an illustration of a simple Bayesian Network. The arrows (directed links) going from one variable to another reflect

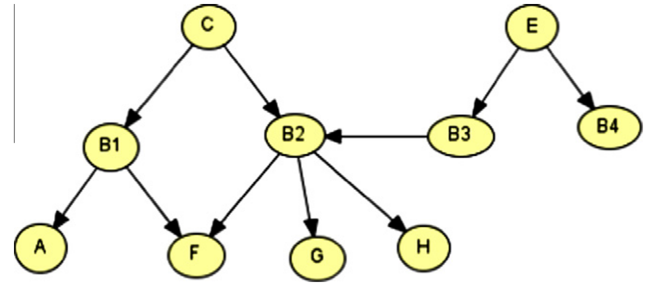

Fig. 1. Bayesian Network example.

the relations between variables. In this example the arrow from C to B2 means that C has a direct influence on B2.

Specifically, a Bayesian Network is a compact and graphical representation of a joint distribution, based on some simplifying assumptions that some variables are conditionally independent of others. As a result the joint probability of a Bayesian Network over the variables $U = \{X_1, \ldots, X_n\}$, represented by the chain rule can be simplified from:

$$P(U) = \prod_i^n P(X_i | x_1, \ldots, x_{i-1})$$

to

$P(U) = \prod_i^n P(X_i = x_i | \text{ parents } (X_i))$, where "parents (X_i)" is the parent set of X_i.

It is this property that makes Bayesian Networks a very powerful tool for representing domains under uncertainty, allowing one to store and compute the joint and marginal distributions more efficiently.

In order to obtain results from Bayesian Networks one does inference. This consists of computing answers to queries made to the Bayesian Network. The two most common types of queries are:

– A priori probability distribution of a variable

$$P(\boldsymbol{A}) = \sum_{X_1} \ldots \sum_{X_k} P(X_1, \ldots, X_k, \boldsymbol{A}) \quad (1)$$

where A is the query-variable and X_1 to X_k are the remaining variables of the network. This type of query can be used during the design phase of a tunnel for example to assess the probability of failure under design conditions (geology, hydrology, etc.).

– Posterior distribution of variables given evidence (observations)

$$P(A|\boldsymbol{e}) = \frac{P(A, \boldsymbol{e})}{\sum_{X_1} \cdots \sum_{X_k} \sum_A P(X_1, \ldots, X_k, \boldsymbol{A}, \boldsymbol{e})} \quad (2)$$

where $\boldsymbol{e}$ is the vector of all the evidence, and A is the query variable and X_1 to X_k are the remaining variables of the network. This type of query is used to update the knowledge of the state of a variable (or variables) when other variables (the evidence variables) are observed. It could be used, for example, to update the probability of failure of a tunnel, after construction has started and new information regarding the geology crossed becomes known.

The most straightforward way to do inference in a Bayesian Network, if efficiency were not an issue, would be to use the equations above to compute the probability of every combination of values and then marginalize out the ones one needed to get a result. This is the simplest but the least efficient way to do inference. There are several algorithms for efficient inference in Bayesian