Risk analysis in tunnel construction with Bayesian networks using mutual information for safety policy decisions

S. GERASSIS[1], Á. SAAVEDRA[2], JULIO F. GARCÍA[1], JOSÉ E. MARTÍN[1] and JAVIER TABOADA[1]
[1]Department of Natural Resources and Environmental Engineering
[2]Department of Statistics and Operational Research
University of Vigo
Lagoas Marcosende, 36310 Vigo, Galicia
SPAIN
sakis@uvigo.es, saavedra@uvigo.es, juliogarcia@uvigo.es, jmartinsuarez@uvigo.es, jtaboada@uvigo.es

Abstract: - Tunnel construction is affected from its origins by different types of uncertainties responsible for innumerable safety risks. This problem has been addressed constantly during the last times achieving positive results, but the complex work scenarios and the common variability of the construction processes prevent putting an end to this problem. For this reason, this study presents an alternative methodology for safety prioritization in tunnel construction gaining relevant information hitherto unknown which can be crucial for policy making in infrastructure projects. The method proposed consists on the Bayesian analysis of data from occupational accidents recorded during the construction of tunnels in the last years. For this purpose, the model variables are rigorously estimated from expert judgement supported by the analysis of data from previous projects. Once the bayesian model is built, the dependencies among the variables are examined using the mutual information. The results obtained from the mutual information analysis allow to detect the main risks responsible for the occurrence of accidents and how they interact. Afterwards, a simplified Bayesian model with the most relevant risk factors affecting safety is built. Through the bayesian inference process, this condensed and validated model facilitates the exploration of significant contributions for safety policy decisions in tunnel construction. Overall, the results obtained provide a deep insight about the most influential factors on which should be focus the efforts to reduce accidents. Several safety risk factors are further influenced by human and organizational factors, whose effect can be reduced in advance. The mechanism of risk migration was better understood when analysing the interaction between the variables in the Bayesian model. In general, the accurate simplification of the model network demonstrated to be a powerful tool to comprehend the uncertainty associated to complex problems.

Key-Words: - Mutual information, supervised learning, occupational accidents, decision making, safety risks.

1 Introduction

Over the last several decades, the underground construction industry has been experiencing a world-wide boom. A look at the tunnel construction statistics since the 1960s shows the obvious consistency of the rising curve with the amount of traffic and civilian infrastructure in these decades [1]. Two main reasons can explain this sudden development. First, the increased necessity of civil tunnels to satisfy the growth of rail and vehicular road traffic. Second, a higher demand for utility tunnels required to create facilities for electricity, water, sewage or modern communication systems.

The continued escalation of the productive requirements in tunnel projects occurs at the same time that higher levels of quality, budget and time completion are required. Tunnel boring machines (TBM) and associated back-up systems are used to highly automate the entire tunneling process, reducing tunneling costs. However, the frequent complex scenarios and the diversity of operations (Fig.1) become occupational accidents a major problem in tunnel construction [2]. The occurrence of accidents are responsible for big overruns and important delays, decreasing considerably the odds of success. Moreover, there exists a generalized public concern about the actual safety of tunnel construction, especially when it comes to coping with uncertainties and their enormous hidden dangers for workers.

All new tunnels now have to be constructed in accordance with the most stringent safety standards, and existing tunnels are being successively upgraded. Safety management plays a fundamental role in order to identify the main risk factors involved in the principal safety violations. In the recent years, there has been an increasing number of studies introducing risk-based analysis into safety control. Qualitative and quantitative risk analysis are the principal risk-based methods used for managing risks in construction projects [3]. They include a wide range of techniques, such as Fault Tree Analysis (FTA), Comprehensive Fuzzy Evaluation Method (CFEM), Influence Diagrams (IDs) or Neural Networks (NNs), among others. These techniques have made a relevant contribution to quality and safety management in complex engineering projects [4-6].

Fig. 1: Complex supporting operation under ground water conditions

However, when dealing with complex scenarios problems arise inevitably due to the common limitation in explicitly representing dependencies of events, updating probabilities and coping with uncertainties. The construction process is affected by different types of uncertainties. Generally, they can be distinguish between the common variability of the construction process and the uncertainty on occurrence of extraordinary events, also denoted as failures of the construction process. Bayesian networks emerged as an efficient solution to address this problem. Their ability to model knowledge, make inferences and reduce uncertainty providing highly visual outcomes caught the attention of many complex engineering projects for safety control [7-9].

The principal objective of this article is trying to identify and cope with occupational hazards in tunnel construction by creating a prioritization of the main safety risk factors that cause the accidents. To better understand the dependencies among the random variables and the type of accident, a Bayesian network (BN) model was built and evaluated with the mutual information (MI) [10] trying to measure the mutual dependence between the type of accident and every random variable. This approach is intended to reduce the problem complexity, obtaining a classification with the risk factors that have a greater influence in the occurrence of accidents.

Subsequently, a new simplified BN model is designed to determine some revealing safety control measures that can then be proposed in advance for risk reduction before an accident occurs. The incorporation of the mutual information to analyse the information exchange among the variables can definitely suppose a new dynamic for safety control in complex engineering environments. The relation of dependence between each variable and the occurrence of the accident could be described as never before. Decision makers and safety engineers would be able to create more accurate policies and construction strategies that concentrate their efforts on human factors and not only on technical issues.

2 Risk factors determination

A good BN modelling practice entails clear definition of the model objective and scope, as well as reliable documentation, data and information sources. A complete revision and identification of possible risk factors was carried out using a 6-year database of accidents in the construction of civil tunnels in Spain.

2.1 Database creation

The study is based on a total of 212 occupational accidents that occurred between 2009 and 2015 during tunnel construction. The database was created from accident reports supplied by the companies operating under the Spanish legislative framework. However, in order to obtain a more calibrated model, further data were collected via questionnaires and personal interviews with relevant company employees. Moreover, it was also made contact with the prevention technicians responsible for accident investigation and safety policies design. They, along with the workers, provided fundamental information regarding the variables that need to be defined in relation to tunnel construction accidents.

Expert judgement is frequently used in developing BNs to generate the model structure and estimate the prior conditional probabilities. The use of expert knowledge is particularly advantageous when empirical data are limited or difficult to obtain. However, modellers should be aware of the uncertainty of data elicited from experts, as humans are susceptible to cognitive biases in judgement and decision making. For this reason, for reliable predictions, the expert estimates were supported by analysis of data from different underground projects.

2.2 Safety risk factors

The revision of the information obtained during the database creation was used to define a set of variables reflecting the full range of the causes that lead to accidents. Due to the complexities in underground construction two types of influential variables concerning the safety issues on tunnel construction are presented (general variables and specific variables) (Table I). In addition to the safety risk factors, the principal types of accidents suffered by workers were registered as well (type of accident). This approach allows the future BN model make inferences about how some specific risk factors influence the occurrence of certain accidents.

2.2.1 General variables

These variables are related as common causes of accidents broadly mentioned in the literature [11-14]. Such variables as *shift work* (V_1), *operator training* (V_6), *operator experience* (V_7) or *order and cleanliness* (V_8) are variables commonly used to illustrate the working conditions and the workers competence for the work to be carried out in the specific context under study.

The definition of these general variables has a positive aspect in regard to possible comparisons between different underground projects. Technical variables depending on the project properties are more difficult to compare, intrinsic to the specific works.

2.2.2 Specific variables in tunnel construction

There are a set of variables that represent certain conditions or activities associated exclusively to tunneling construction, which can have a big influence on the occurrence of accidents. Some of these variables refer to geological and design factors, such as *tunnel length (V10)*, *state of the floor surface (V14)* or *excavation method* (V_{17}). Others emphasize the importance of stability and tunnel support, for example, *stability of the excavation front* (V_{15}), *rock bolts installation (V29)* or *tunnel waterproofing* (V_{33}). Finally, there are some variables related to technical and operational aspects including, *outsourcing in the same cut (V37)* or *works execution deadline (V38)*.

Table I. Risk factors identified in tunnel construction

General variables		Specific variables	
V_1	Shift work	V_{10}	Tunnel length
V_2	Shift duration	V_{11}	Tunnel section
V_3	Machinery age	V_{12}	Type of excavation section
V_4	Machinery maintenance	V_{13}	State of haul roads
V_5	Operator seniority	V_{14}	State of the floor surface
V_6	Operator training	V_{15}	Stability of the excavation front
V_7	Operator experience	V_{16}	Simultaneity of operations
V_8	Order and cleanliness	V_{17}	Excavation method
V_9	Signposting and signalling	V_{18}	Loading machinery
		V_{19}	Conveyor belt
		V_{20}	Type of hauling vehicle
		V_{21}	Vehicles interference
		V_{22}	Water drainage
		V_{23}	Walls stability
		V_{24}	Drilling machinery
		V_{25}	Dust collection
		V_{26}	Gas detection
		V_{27}	Tunnel profiling
		V_{28}	Tunnel initial support
		V_{29}	Steel ribs installation
		V_{30}	Rock bolts installation
		V_{31}	Type of shotcrete
		V_{32}	Tunnel lining
		V_{33}	Tunnel waterproofing
		V_{34}	Ventilation
		V_{35}	Lighting
		V_{36}	Outside communication systems
		V_{37}	Outsourcing in the same cut
		V_{38}	Works execution deadline
		V_{39}	Operator assigned work

2.2.3 Type of accident

As a further variable, in addition to the 39 in Table 1, the type of accident states the typology of the accident suffered by workers during the tunnel construction. This variable is of paramount importance. Depending on the type of accident a worker suffers, certain variables are going to have a greater or lesser influence. The definition of the principal typologies found were as follows:

- Falls from the same or different height.
- Detachment or handling-induced falls of loose objects.
- Collision or hit with objects.
- Projection of fragments and particles.

- Entrapment by or between objects.
- Overexertion.
- Exposure to corrosive substances or electrical contacts.
- Fires.
- Accidents involving vehicles.

3 Bayesian networks and mutual information in safety risk analysis

In machine learning and data mining, dimensionality reduction refers to the process of reducing the number of random variables under consideration, obtaining a subset of principal variables, typically from large databases [15]. In this case, the database created with accidents recorded in tunnel construction it was used in a first step to create a general Bayesian model. This model was afterwards reduced through the analysis of the mutual information between the variables.

The topological structure of a Bayesian model reflects the variables' dependency and describes the probability distribution of certain events occurring in specific conditions. If $X = \{X_1, X_2, ..., X_n\}$ is a set of *m*-dimensional variables, then a Bayesian network is formally defined as a couplet $X = \langle G, P \rangle$ where *G* is a directed acyclic graph in which each node represents one of the variables $X_1, X_2, ..., X_n$ and each arc represents a direct dependency relationship between these variables; and *P* is a set of parameters that quantifies the network, containing the probabilities for each possible value x_i for each variable X_i.

From the decomposition theorem, the joint probability P, under the hypothesis that each node is independent of its non-descendants, can be calculated. Therefore, the Bayesian network has a single joint probability distribution given by:

$$P(X) = P(X_1, X_2, ..., X_n) = \prod_{i=1}^{n} P(X_i | X_{j(i)}) \quad (1)$$

where $X_{j(i)}$ is the set of parent variables of X_i for direct acyclic graph *G*. Consequently, the application of Bayes' theorem enables to determine the posterior probability of the variable of interest through the inference process.

When building a Bayesian network, it is necessary to explore different structures [16]. Firstly, a supervised learning approach was used for generating a model for the target variable's prediction autonomously from data. In this approach, the only guidance provided is the node of interest, the *type of accident*, representing the target variable for the machine learning process.

For computation and data analysis, Artificial Intelligence software BayesiaLab v6.0.7 [17] was used. A direct network with a Naive Bayes algorithm was built in order to minimize the network complexity, making it easier for the computation of the mutual information. The concept of mutual information is intricately linked to that of entropy of a variable, a fundamental notion in information theory that defines the amount of information held by the variable itself.

Shannon Entropy [18-19] was used for the computation of information exchanged between the target variable and every risk factor (Table I). The definition of Shannon Entropy, of a discrete variable X is:

$$H(X) = -\sum_{x \in X} p(x) \log_2 p(x) \quad (2)$$

The difference between the marginal entropy of the target variable and the conditional entropy of a given target is formally known as mutual information (MI). More generally, the mutual information between two variables X and Y is defined by [20]:

$$MI(X,Y) = H(X) - H(X | Y) \quad (3)$$

which is equivalent to:

$$MI(X,Y) = \sum_{x \in X} \sum_{y \in X} p(x,y) \log_2 \frac{p(x,y)}{p(x)p(y)} \quad (4)$$

The computation of the mutual information, between the *type of accident* and each variable, in the form of risk factors, is represented by the Bayesian probability allocated to each class of the target variable. Thus, the predictors providing the maximum information can be properly identified, highlighting their predictive importance and repercussion as risk factors in the occurrence of accidents in tunnel construction

4 Simplified Bayesian model

In a second step, once the main risk factors on which efforts to reduce accidents should be focus were identified, a simplified Bayesian model was carried out as a support tool in safety studies conducted in the planning stage of tunnel construction. On this occasion, the supervised learning during the network construction was evaluated in order to obtain a BN model with a complexity determined by the adequate amount of relationships among variables representing a common field work. The Minimum Description Length (MDL) [21], in which BayesiaLab's algorithms are based on, was used for measuring the quality of the network with respect to the underlying data. This strategy is now feasible and highly practical, because since the problem dimensionality is reduced more significant cause and effect relationships can be discovered in a much more efficient way for safety analysis.

The MDL score is composed of two parts: one to score the structure and other one to quantify how well the network fits the data. The MDL must be minimized to obtain the best solution that leads to the best compression of the data. Formally, can be written as [20]:

$$MDL(B,D) = \propto DL(B) + DL(D|B) \qquad (5)$$

- $\propto$ represents the structural network coefficient (SC). This parameter, which default value is 1, permits changing the network complexity. The lower the value of α, the greater the complexity of the resulting network.
- $DL(B)$ is the number of bits to represent the Bayesian network, *B,* graph and probabilities. For this term, the minimum value is obtained with the simplest network structure.
- $DL(D|B)$ is the number of bits to represent the dataset, *D,* given the Bayesian network, *B.* Here, the minimum value corresponds to the fully connected network, in which no structural independencies are stated.

Thus, minimizing this score consists in finding the best trade-off between both terms. This can be achieved finding a point in between the simplest structure where the network is fully unconnected and the fully connected network, in which no structural independences are stated [20].

Finally, when the suitable model is found, the BN can be used to exploit the conceptual field by applying intercausal reasoning [22]. Major accidents generally do not originate from a single circumstance [23]. Their occurrence is determined by the interaction between the physical construction processes, the behavior of operators and the organizational elements.

The intercausal reasoning technique opens the door to explore the probability of occurrence of

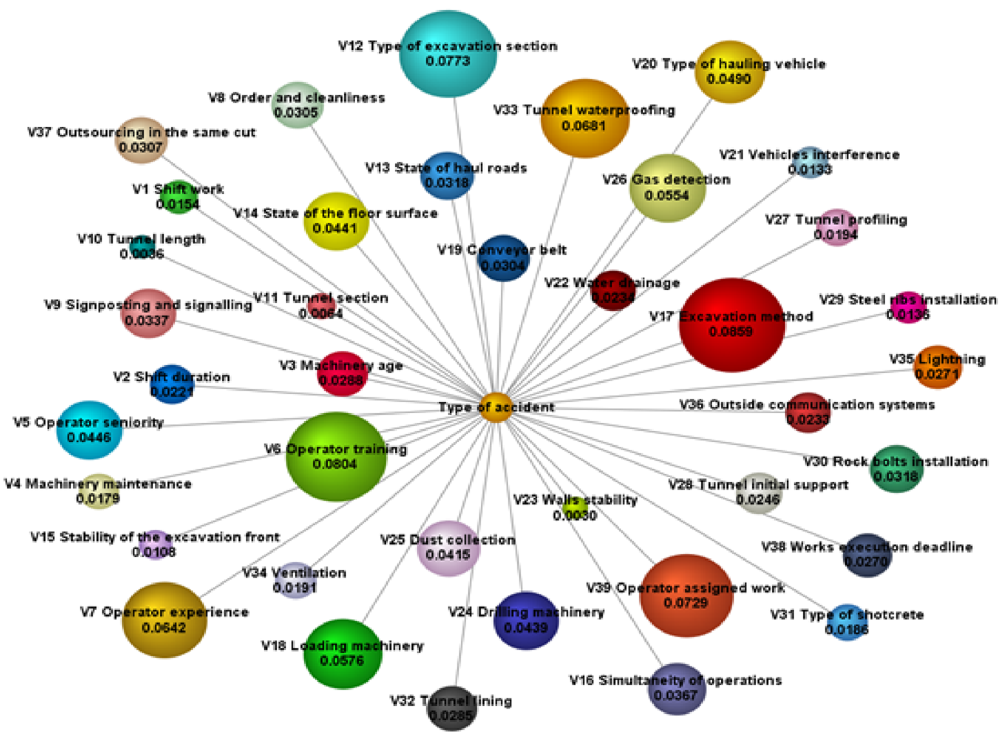

Fig.2: Supervised BN with Naive Bayes algorithm. The mutual information result for each node is represented in a proportional way to its size.

certain events and its relation with producing particular types of accidents. The inference results will allow to define safety strategies which address specifically the root causes responsible for every type of accident in tunneling construction.

5 Results and discussion

The full range of risk factors identified in tunnel construction were ordered using the mutual information with regard to the type of accident. A simplified Bayesian model was built and the most likely accidents were quantified and analyzed. The use of intercausal reasoning showed the influence of specific variables in causing accidents. The results obtained are widely described in the next subsections.

5.1 Prioritisation of safety risk factors

The Bayesian analysis executed with the supervised learning approach and the measure of the mutual information allowed to establish a classification of the most influential variables causing accidents. Fig. 2 shows the supervised BN built using the Naive Bayes algorithm. Every node contains the mutual information result. The units of mutual information are bits, because the log base 2 is used in (4). A complete summary of the mutual information result for every variable is presented in Table 2, ordered from the most to the least influential factor causing accidents in tunnel construction.

As can be observed, *excavation method (V_{17}), operator training (V_6), type of excavation section (V_{12}), operator assigned work (V_{39}), tunnel waterproofing (V_{33})* and *operator experience (V_7)* are identified, in that order, the six factors most affecting the occurrence of accidents. Conversely, *steel ribs collocation (V_{29}), vehicles interference (V_{21}), stability of the excavation front (V_{15}), tunnel section (V_{11}), tunnel length (V_{10})* and *walls stability (V_{23})* are the factors that less affect the occurrence of accidents. Some of these results are undoubtedly surprising.

For example, it is an extended belief among civilians that the excavation stability is the major consequence of accidents in tunnel construction. However, the tremendous advances in the last decades in supporting technology of the underground construction industry have reduced notably this circumstance. This can explain that *stability of the excavation front (V_{15})* or *walls stability (V_{23})* are among the factors that affect the least accidents occurrence (Table II).

Table II. Risk factors classification with respect to the information gain brought by each node to the knowledge of the type of accident

	Network Nodes	Mutual Information	Relative significance
V_{17}	Excavation method	0.0859	1.0000
V_6	Operator training	0.0804	0.9368
V_{12}	Type of excavation section	0.0773	0.9003
V_{39}	Operator assigned work	0.0729	0.8486
V_{33}	Tunnel waterproofing	0.0681	0.7930
V_7	Operator experience	0.0642	0.7476
V_{18}	Loading machinery	0.0576	0.6712
V_{26}	Gas detection	0.0554	0.6451
V_{20}	Type of hauling vehicle	0.0490	0.5704
V_5	Operator seniority	0.0446	0.5191
V_{14}	State of the floor surface	0.0441	0.5142
V_{24}	Drilling machinery	0.0439	0.5115
V_{25}	Dust collection	0.0415	0.4836
V_{16}	Simultaneity of operations	0.0367	0.4277
V_9	Signposting and signalling	0.0337	0.3924
V_{30}	Rock bolts installation	0.0318	0.3707
V_{13}	State of haul roads	0.0318	0.3702
V_{37}	Outsourcing in the same cut	0.0307	0.3579
V_8	Order and cleanliness	0.0305	0.3548
V_{19}	Conveyor belt	0.0304	0.3540
V_3	Machinery age	0.0288	0.3359
V_{32}	Tunnel lining	0.0285	0.3323
V_{35}	Lightning	0.0271	0.3156
V_{38}	Works execution deadline	0.0270	0.3143
V_{28}	Tunnel initial support	0.0246	0.2866
V_{22}	Water drainage	0.0234	0.2724
V_{36}	Outside communication system	0.0233	0.2714
V_2	Shift duration	0.0211	0.2575
V_{27}	Tunnel profiling	0.0194	0.2255
V_{34}	Ventilation	0.0191	0.2219
V_{31}	Type of shotcrete	0.0186	0.2164
V_4	Machinery maintenance	0.0179	0.2086
V_1	Shift work	0.0154	0.1788
V_{29}	Steel ribs installation	0.0136	0.1579
V_{21}	Vehicles interference	0.0133	0.1547
V_{15}	Stability of the excavation front	0.0108	0.1256
V_{11}	Tunnel section	0.0064	0.0742
V_{10}	Tunnel length	0.0036	0.0415
V_{23}	Walls stability	0.0030	0.0349

However, the safety risks seem to move towards the increasingly diverse and complex excavation methods *(V_{17})*, which require more operators with an appropriate training in occupational risks to accomplish the assigned works *(V_{39})*, where experience plays a fundamental role *(V_7)*.

These results shed light about the importance of training for workers. The more knowledge a worker has, the less risks will materialize. Therefore, the number of accidents will decrease and works will be completed more frequent at the time scheduled.

5.2 Bayesian model simplification

The central premise of using mutual information as variable selection is that the data contains many features that are either redundant or irrelevant, and thus can be removed without incurring a significant loss of information [24].

The factors that statistically have shown a greater influence on accidents occurrence can be now used to create a simplified BN. This simplification does not invalidate a more generalized preventive study that takes into account all the possible factors. However, once it is demonstrated the low influence of some factors, their use increases the network complexity becoming the model redundant and inaccurate leading to possible deviations between the predicted results and those observed in reality.

For this reason, in the simplified model were only considered those variables with a relative significance of at least 0.30 with respect to the information brought by the variable to the knowledge of the target node (red line, Table II). The relative significance expresses the ratio between the mutual information brought by each variable and the maximum mutual information brought by any variable, which corresponds to *excavation method (V17)*. This choice allows to reduce the model from 39 to 24 risk factors. The new BN model is shown in Fig. 3.

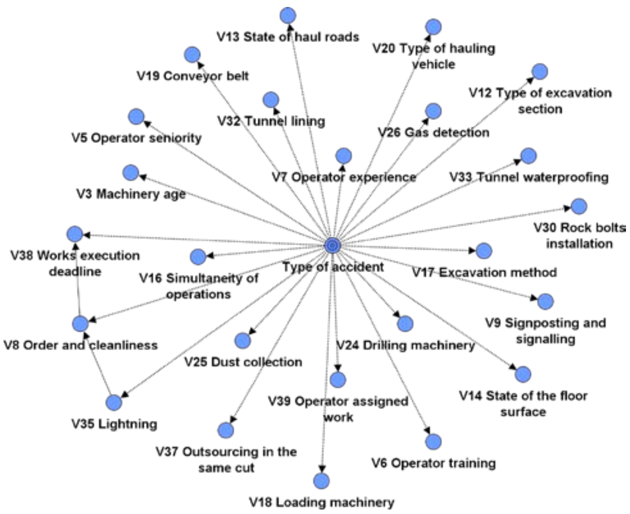

Fig. 3: Simplified Bayesian network with SC=1

This network was built using an Augmented Naive Bayes algorithm with a SC value of 1, which is the default value set to BayesiaLab's algorithms. The possibility of manipulating the SC makes it possible for the analyst to change the weight of the structural part in the MDL score. The possible range of values for this SC parameter is 0 to 150 [20]. An SC value of 0 means that the MDL score is exclusive based on data fit, thus potentially resulting in a fully-connected network. At the other extreme, an SC value of 150 favors simplest and often unconnected structures. As such, this parameter works as a threshold.

The higher the SC value, the stronger the probabilistic relations would have to be to result in a corresponding arc in the network. Conversely, the lower the SC value, the weaker the probabilistic relation can be while still being represented with an arc. The key point it is to find the right level of complexity of the network.

Lowering the SC value can be particularly useful if the number of available observations is not excessively big, like in this case (212 accidents). However, choosing too low of an SC value might result in learning insignificant relationships and thus overfitting the network model to the data. To achieve a right balance it was performed a SC analysis in order to examine the data-to-structure ratio as a function of the SC (Fig. 3).

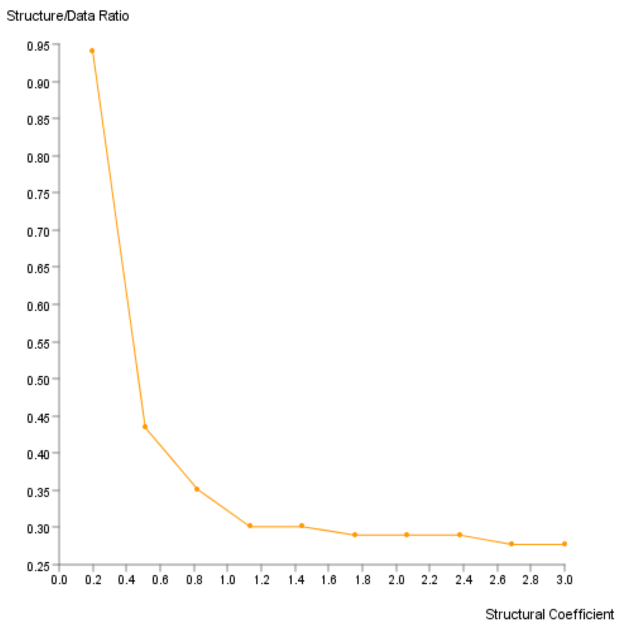

Fig. 4: Data-to-structure ratio as a function of the SC value

By analyzing the graph in Fig. 4, moving from right to left along the x-axis, an important inflection point of the curve can be seen around SC = 0.50. Below that value, the structural complexity of the network is increasing faster than the data likelihood. For this reason, it is chosen SC = 0.50 and relearn the network on that basis with the supervised learning algorithm Augmented Naive Bayes.

The resulting network in Fig. 5 is now more sophisticated than the original simplified network in Fig. 3, in terms of arcs representing dependencies of events. At this point, the factors whose cause-effect relationships are shown can be considered the most relevant for safety management in tunnel construction, according to the available data.

Fig. 5. Simplified Bayesian network with SC=0.50

Additionally, once the final BN was built (Fig. 5) the following results with regard to the type of accident were obtained (Fig. 6). The most likely type of accident was *collision or hit with objects (28.30%)*, closely followed by *falls from the same or different height (21.70%)* and *overexertion (21.23%)*.

The least likely accidents were *fires (0.94%)*, *projection of fragments or particles (1.42%)* and *entrapment by or between objects (4.25%)*. Other interesting results concerning safety in tunnel construction were found looking at some specific factors. For example, at the moment of an accident, a total of *83%* of the *works execution deadline* (V_{38}) was delayed and tunnel *lightning* (V_{35}) was unsatisfactory the *65%* of the time. Another key factor in occupational accidents is the *order and cleanliness* (V_8), which was found deficient in *82%* of cases. Regarding technical aspects, the *excavation method* (V_{17}) most likely to be associated with an accident was drilling and use of explosives *(63%)*, far more than those excavation methods executed with excavators or roadheaders *(37%)*.

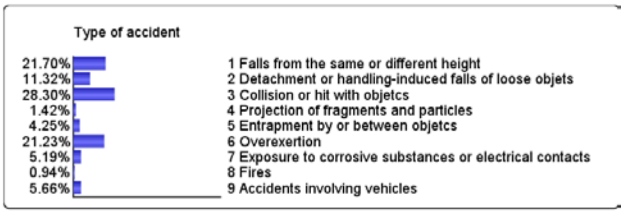

Fig. 6. Probability of occurrence for every type of accident

5.3 Inference results

One of the inherent abilities of Bayesian networks to explicitly model uncertainty is that they compute inference omni-directionally [22]. Given an observation with any type of evidence on any of the networks' nodes, it is possible to compute the posterior probability of all other nodes in the network, regardless of arc direction. The impact of lack of experience joined to the lack of training is corroborated as a common cause of occupational accidents by several studies [25-29]. In Fig. 6 is evaluated this circumstance setting a hard evidence where there is no uncertainty regarding the state of the variable (V_7), *P(Operator experience=No)=100%* .

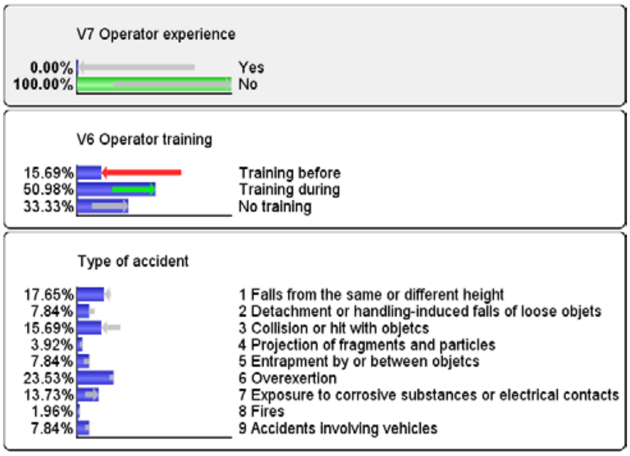

Fig. 7. Operator experience evidence-setting and inference results for operator training and type of accident.

The inference results in Fig. 6 show that the operator inexperience is linked to high probabilities of an inadequate training, due to only 15.69% of operators had been trained before the commencement of the activity. Moreover, the lack of experience increases the risk of suffering certain accidents, such as those related to *overexertion* (from 21.23% to 23.53%) or *exposure to corrosive substances or electrical contacts* (from 5.19% to 13.73%). In order to have a clearer picture of lack of experience coupled with the lack of training, the inference results for every type of accident were plotted in Fig. 7.

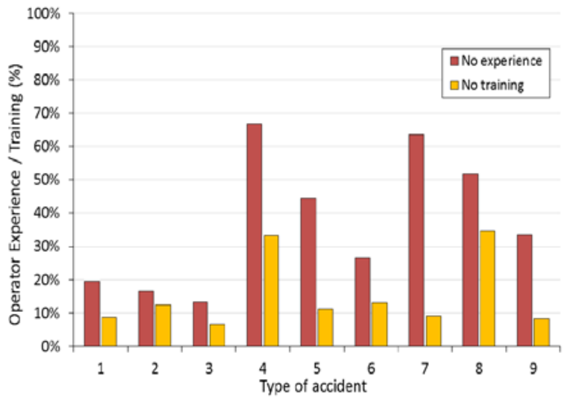

Fig. 8. Inexperience and lack of training impact on the different types of accidents in tunnel construction

Accidents caused by *projection of fragments and particles* (type 4) and *exposure to corrosive substances or electrical contacts* (type 7) are the most sensitive to operators' inexperience. When it comes to lack of training, *projection of fragments and particles* (type 4) and *fires* (type 8) are the accidents with the highest risk. This seems reasonable having into account that compliance with safety protocols it is something crucial when there is a fire or during the use of explosives. The ignorance of operators about how to react to unexpected situations promotes safety violations many related to a deficient hazard perception. Conversely, it is noticeable that types 1, 2 and 3, due to falls, detachments and collisions are the accidents for which operators present the lowest levels of influence regarding inexperience and lack of training.

In summary, this knowledge modelling process and reasoning under uncertainty demonstrates the potential for safety risk analysis of reducing the problem complexity by using Bayesian networks.

6 Conclusions

Safety management of tunnel construction deals with complex work scenarios where accidents occur owing to a wide range of risk factors difficult to assess. This paper shows Bayesian networks and the use of mutual information as a powerful tool for data integration and knowledge reasoning in order to identify and cope with the principal occupational hazards responsible for accidents, as well as overruns and delays affecting the underground construction industry.

A full range of risk factors identified in tunnel construction were prioritized using the mutual information with regard to their influence on accidents occurrence. The *excavation method* and the *type of excavation section* together with several issues concerning operators, such as *training*, *experience* or the *assigned work* were found the most influential factors on which efforts to reduce accidents should be focused.

From these factors, a simplified Bayesian model was built, paying special attention to achieve a network design that accurately represent the dependencies of events. The resulting model allowed to identify *collision or hit with objects* the most likely type of accident, followed by *overexertion*. Finally, exploiting the possibility of Bayesian networks to analyse the conceptual field by applying intercausal reasoning different conclusions concerning the operators' inexperience and lack of training were brought to light. The results obtained suggest that this type of network prioritisation and simplification can be reliably employed in the future in other fields with complex scenarios.

References:

[1] K. S. Lane, Tunnels and underground excavations, Encyclopædia Britannica (2017, May 10) [Online]. Available: https://www.britannica.com/technology/tunnel/Future-trends-in-underground-construction

[2] B. J. Arends, S. N. Jonkman, J.K. Vrijling, P. van Gelder, Evaluation of tunnel safety: towards an economic safety optimum, *Reliab. Eng. Syst. Saf.* Vol. 90, No. 2, 2005, pp.217-228.

[3] L. Zhang, X. Wu, L. Ding, M. J. Skibniewski, Y. Yan, Decision support analysis for safety control in complex project environments based on Bayesian Networks, *Expert Syst. Appl.* Vol. 40, 2013, pp.4273-4282.

[4] V. Carr, J. H. M. Tah, A fuzzy approach to construction project risk assessment and analysis: Construction project risk management system, *Adv. Eng. Softw.* Vol. 32, 2001, pp.847-857.

[5] M. Correa, C. Bielza, J. P. Teixeira, Comparison of Bayesian networks and artificial networks for quality detection in a machining process, *Expert Syst. Appl.* Vol. 36, 2009, pp.7270-7279.

[6] F. Piniella, A. Fern, M. A. Ndez-Engo, Towards system for the management of safety on board artisanal fishing vessels: Proposal for checklists and their application, *Saf. Sci.* Vol. 47, 2009, pp.265-276.

[7] S. Gerassis, J. E. Martin, J. F. García, Á. Saavedra, J. Taboada, Bayesian decision tool for the analysis of occupational accidents in the construction of embankments, *J. Constr. Eng. Manage.* Vol. 143, No. 2, 2017, pp. 273-282.

[8] R. L. Sousa, H. H. Einstein, Risk analysis during tunnel construction using Bayesian networks: Porto metro case study, *Tunn. Undergr. Space Technol.* Vol. 27, 2012, pp.86-100.

[9] S. Leu, C-M. Chang, Bayesian-network-based safety risk assessment for steel construction projects, *Accid. Anal. Prev.* Vol. 54, 2013, pp.122-133.

[10] T. M. Cover, J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons Inc., 1991.

[11] A. M. Williamson, A-M. Feyer, Causes of accidents and the time of day, *Work Stress* Vol. 9, 1995, pp. 158-164.

[12] J. E. Martín, T. Rivas, J. M. Matías, J. Taboada, A. Argüelles, A Bayesian network analysis of workplace accidents caused by falls from a height, *Saf. Sci*. Vol.47, No.2, 2009, pp. 206-214.

[13] S. Salmien, T. Tallberg, "Human errors in fatal and serious occupational accidents in Finland," *Ergonomics*. Vol. 39, 1996, pp. 980-988.

[14] A. Oliver, A. Cheyne, J. M. Tomás, S. Cox, The effects of organizational and individual factors on occupational accidents, *J. Occup. Organ. Psychol*. Vol. 75, 2002, pp. 473-448.

[15] S. T. Roweis, L. K. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science. Vol. 290, 2000, pp. 2323-2326.

[16] P. M. Kuhnert, K. R. Hayes, How believable is your BBN? *18th World IMACS/MODISM Congress*. Cairns, Australia, 2009.

[17] Bayesia S.A.S. BayesiaLab Software. Parc Cérès, Bâtiment Nº21, rue Ferdinand Buisson, 53810 Changé, France, 2017. http://www.bayesia.com/

[18] M. Borda, Fundamentals in Information Theory and Coding, Springer pp.11, 2011.

[19] C. E. Shannon, "A mathematical theory of communication," Bell System Technical Journal. Vol. 27, 1948, pp.379-423.

[20] S. Conrady, L. Jouffe, Bayesian Networks & BayesiaLab - A Practical Introduction for Researches, Bayesia USA: Franklin, TN., 2015.

[21] J. Rissanen, Modeling by shortest data description, *Automatica*. Vol. 14, 1978, pp. 465-658.

[22] B. D'Ambrosio, Inference in Bayesian networks, *AI Magazine*. Vol. 20, No. 2, 1999, pp. 22-36.

[23] J. Reason, Managing the risks of organizational accidents, Ashgate, Bookfield, VT, USA, 1997.

[24] H. Peng, F. Long, C. Ding, Feature selection based on mutual information criteria of max-dependency, max-relevance and min-redundancy, *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Vol. 27, No. 8, 2005, pp. 1226–1238.

[25] P. S. Paul, J. Maiti, The role of behaviour factors on safety management in underground mines, *Saf. Sci*. Vol. 45, 2007, pp.449-471.

[26] M. T. Birgonul, I. Dikmen, C. Budayan, T. Demirel, An expert system for the quantification of fault rates in construction fall accidents, *Int. J. Occup. Saf. Ergon.* Vol. 22, 2016, pp.20-31.

[27] S. Leitao, B. A. Greiner, Organisational safety climate and occupational accidents and injuries: an epidemiology-based systematic review, *Work Stress*. Vol. 30, 2016, pp.71-90.

[28] V. Gharibi, S. B. Mortazavi, A. J. Jafari, J. Malakouti, M. Bagheri, H. Abadi, The relationship between workers' attitude towards safety and occupational accidents experience, *Int. J. Occup. Hyg*. Vol. 8, No. 3, 2016, pp.145-150.

[29] J. F. García, J. E. Martín, S. Gerassis, A. Saavedra, J. Taboada García, Bayesian analysis of risk associated with workplace accidents in earthmoving operations, *Informes de la Construcción*, Vol. 69, No. 546, 2017.