Towards A Model-Based Asset Deterioration Framework Represented by Probabilistic Relational Models

Haoyuan Zhang & D. William R. Marsh

Risk and Information Management Research Group, School of Electronic Engineering and Computer Science, Queen Mary University of London

ABSTRACT: Most asset deterioration tools are designed for a specific application, as a consequence, a small change of the specification may result in a complete change of the tool. Inspired by the model-based approach of separating problem specification from analysis technique, we propose a model-based asset deterioration assessment framework using probabilistic relational models. The probabilistic relational models express abstract probabilistic dependency covers a range of deterioration modelling assumptions. An expert in the domain of asset deterioration can then use his knowledge of the factors that affect deterioration to customise the abstract models to a specific application, without requiring a detailed understanding the underlying computational framework. We illustrate the use of the framework with multiple variants of deterioration models.

1 INTRODUCTION

Traditionally, inspection and maintenance of infrastructure has followed a fixed time interval. One idea to make inspection more cost-effective is to use a statistical model to predict the rate of asset deterioration and use the predictions to plan detailed inspections or maintenance. A range of deterioration models has been developed in different field, from railway track (Guler et al., 2011) to bridge (Sobanjo, 2011). Despite having common objectives such as condition prediction and using the prediction to leverage maintenance planning, the models differ in many ways. For example, the deterioration distribution and the grading system for asset condition may differs depending on the asset type or the standards set by inspection agencies. Our aim is to build a unified framework that is general enough to encode a wide variety of deterioration models.

The approach of providing unified tools, so called model-based system engineering, has been advocated in both industry and academia. It aims to provide descriptive modelling of systems, common to different system analysis techniques. Importantly, this approach aims to enable decision makers to use analyses without a detailed knowledge of the underlying mathematical models. We review previous work in Section 2 and consider how it applies to deterioration models.

In Section 3, we describe a framework for maintenance domain experts, which does not require a detailed understanding of the underlying deterioration models. Our framework extends standard hierarchical Bayesian models with relational schema, allowing model variants to be expressed using domain concepts. In Section 4, we illustrate the use of the framework with a variety of deterioration model.

2 MODEL-BASED APPROACH

The emerging field of model-based system engineering focuses on bridging the gap between problem specification and modelling (Estefan, 2007), with an agile modelling formalism to meet different modelling requirements without changing the entire tool (Prosvirnova et al., 2013). The model-based approach formalises the system development process using a unified language (e.g. SysML language) to provide a platform integrating different modelling approaches and system analysis methods. This formalism has been extended to the safety and reliability domain, so called model-based safety assessment (MBSA) (see Lisagor et al. (2011)). MBSA aims to unify classical safety and reliability modelling methods (e.g. fault tree and stochastic process), and to generate an integrated structure for a range of safety and reliability analysis (e.g. fault tree analysis and system diagnosis). For example, the AltaRica modelling language (Arnold et al., 1999), separates system specification from analysis with a range of reusable techniques.

The MBSA concept has been applied in deterioration assessment in recent years, with the modelling of stochastic process and Markov chain for complex system developed in project AltaRica 3.0 (Prosvirnova et al., 2013). However, to our knowledge, the current practice of MBSA does not yet encompass learning from data, which is a component of deterioration modelling when since we wish to learn deterioration rates from inspection data. Fortunately, advances in machine learning provide a promising perspective for tackling these problems.

Previous studies have shown the power of a hierarchical Bayesian network (BN) based approach to learn asset deterioration rates from data and how it
can be adapted when there is insufficient data, both with expert knowledge (Frangopol et al., 2004, Zhang and Marsh, 2018), or by learning from similar groups (Memarzadeh et al., 2016, Zhang and Marsh, 2018). In the work of Zhang and Marsh (2018), six generic BN models for asset deterioration were developed, which both provides us the possibility of adopt different deterioration models, but also enables us to include alternative data and unused expert knowledge. These model variants cannot yet be presented to an asset deterioration domain expert in a unified framework: adapting the underlying concepts to a particular context requires a deep understanding of their implementation as BNs.

In model-based machine learning (MBML) (see Bishop (2013) and Ghahramani (2015)), models and problem specifications are defined in a compact language, while inference or machine learning algorithm codes are generated automatically. Bayesian networks are such a language, though they lack structure. More recently, probabilistic programming languages such as Figaro (Pfeffer, 2009), has been developed which could also be used in our framework.

Model-based approaches, both in MBML and MBSA, often use the object-oriented paradigm to provide a library of generalized models for reuse. This is not provided by traditional BNs, with a fixed set of variables and relationships. This issue has been widely researched for BNs, with proposals including idioms in Neil et al. (2000) and fragments in Laskey and Mahoney (2000). Probabilistic relational models (PRM), developed by Koller (1999) combines relational structure with probabilistic graphical models (i.e. BNs). A PRM combines probabilistic dependencies with a relational schema that describes the entities in the problem domain. This representation provides a separation between model library and structure relationships.

Therefore, we propose to develop a model-based framework for asset deterioration assessment in the spirit of MBSA. The framework separates reusable low-level models from modelling choices and asset descriptions. The framework is encoded with a PRM representation of a hierarchical Bayesian network, with a range of generalised models for asset deterioration each represented by its probabilistic dependencies, and the problem specification of the target domain is represented as the relational schema.

3 MODEL-BASED ASSET DETERIORATION ASSESSMENT FRAMEWORK

3.1 Asset Deterioration Model using Hierarchical BNs

3.1.1 A Simple Deterioration Model

For a system that is either working or failed, given historical data on the times that it remained in working condition, we can estimate the distribution of time for its transition to the failed state and so predict its likelihood of failing. A basic deterioration BN model, from Zhang and Marsh (2016), is shown in Figure 1. This is a hierarchical BN model that both learns from data and can be used for decision support.

However, this specific model can only be used to describe a type of asset with two-state and deterioration that follows a one-parameter distribution. This is not usually the case in asset deterioration, for example a four point grading system is used to describe bridge condition, and a two-parameter Weibull distribution is used to fit the bridge transition distribution in Sobanje (2011). So instead, the model has to be adapted: the variables are similar but the number of them and links between them must change.

![Figure 1. A simple deterioration model.](image)

For the system that can be in either working or failed state, the distribution of times that it remained in working condition, we can estimate the distribution from the historical transition data and predict the likelihood of failing. A basic deterioration BN model, from Zhang and Marsh (2016), is shown in Figure 1. This is a hierarchical BN model that both learns from data and can be used for decision support.

However, this specific model can only be used to describe a type of asset with two-state and deterioration that follows a one-parameter distribution. This is not usually the case in asset deterioration, for example a four point grading system is used to describe bridge condition, and a two-parameter Weibull distribution is used to fit the bridge transition distribution in Sobanje (2011). So instead, the model has to be adapted: the variables are similar but the number of them and links between them must change.

![Figure 2. Effects of prior knowledge and data quantity in distribution training.](image)

The structure of the model may also need to change when we have limited domain data. Although the parameters of the transition distribution are learnt from the historical transition data, some prior knowledge of the transition distribution is also required. In Figure 2, two different prior probability distributions have been used: i) an uninformative prior and ii) an informative prior, available when there is good knowledge representation of the deterioration. Figure 2(a) shows that with good prior, we can provide a good estimate of the true distribution with only a little data, while Figure 2(b) shows a...
larger dataset will give a correct estimate of the parameter even the prior is weak.

However, when failure data is scarce (which is the usual case in slow deteriorating asset, for example, bridges) or knowledge is poor for a particular asset class (which is also usual for new assets), we can combine data from asset types that, though not identical, are similar and so are believed to have the similar deterioration rate (Morcos, 2011). This kind of approximation is necessary, especially for assets types that are inspected infrequently so that the deterioration dataset is not large enough for each asset types. Two techniques are proposed in Zhang and Marsh (2016): one is to add another layer of parameter (hyper-parameter) to form a hierarchical BN that can group or pool data from different asset types sufficiently to overcome an uninformative prior; the other technique is to use influencing factors to adjust the transition distribution of a specific asset from distribution learned from a pool of similar assets (i.e. assets of the same type). Both methods may change the BN’s structure depending on how assets are assigned to groups and what other factors influence the transition times. We refer to these (and related) issues as ‘modelling assumptions’.

3.1.2 A Framework for Expressing Modelling Assumptions

To address the problem of many variants of the deterioration models, we provide a framework to help domain experts express modelling assumptions. The stages of model-based asset deterioration assessment framework can be illustrated in Figure 3:

- The model library encodes the possible dependencies of probabilistic models in the problem domain.
- Model selection uses modelling assumptions to determines what models, knowledge and data are included in the problem model.
- The relational database includes the configuration and failure data; its schema is derived from the modelling assumptions.
- Instantiation and inference, performed automatically, are used to evaluate queries on the model for domain decision support.

The following sections describe each aspects of the framework.

3.2 Model library: Abstract PRM

The generic models in the model library are represented as abstract probabilistic relational models (PRMs). Figure 4 shows an example developed from Zhang and Marsh (2018).

![Figure 4. Probabilistic dependencies represented as an abstract PRM](image)

In Figure 4, a square (called a class) defines a group of identical objects that share the same set of variables or probabilistic models. An oval defines a variable, and directed edge defines the dependency of variables. Aggregation is defined by a bold arrow. *N* represents a fixed multiple relationship, and * represents a multiple relationship of uncertain degree. The classes are as follows:

<table>
<thead>
<tr>
<th>Class</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Asset</strong></td>
<td>We wish to predict the state of a specific asset, conditioned on its previous inspection and the deterioration data of similar assets. This prediction will inform decisions about maintenance and inspection.</td>
</tr>
<tr>
<td><strong>Transition</strong></td>
<td>Objects of this class represent transitions in a Markov chain, where the conditional probability of moving into future state $S_{t+1}$ at time $t+1$ given the present state $S_t$ at time $t$ follows a distribution with parameters learnt from data.</td>
</tr>
<tr>
<td><strong>Data</strong></td>
<td>Data is gathered from inspections, each giving information about the current state of an asset. Different types of data are used: for example, it is common to have only censored data giving a time after which a transition occurred. This is modelled as constraints on the transition time.</td>
</tr>
<tr>
<td><strong>Group</strong></td>
<td>Assets of the same type form a group. A group is represented by the model by parameters of the</td>
</tr>
</tbody>
</table>
distribution (for each transition) learnt from historical data for this type of assets. Since the parameters are learnt, their values are uncertain and the model include them as probabilistic variables.

| Parameter | The population as a whole also has distribution parameters. The similarity of each group of assets to the population as a whole is judged and this establishes a way to learn a group’s distribution parameters from data of other closely related groups. |
| Factor | The idea of assets of the ‘same type’ is defined in relation to properties of the asset that influence deterioration rate. However, if all relevant factors are used to distinguish groups then there is likely to be insufficient historical data to estimate the transition distribution parameters. Therefore, groups need to be defined by the factors that are most important (and are known for all assets). Other factors – for example, loading and environment condition – can be used to estimate a target asset’s distribution. |

We describe this model as ‘abstract’ as it represents number of different structures. In particular, the following issues need to be resolved to give a specific model:

- The distribution (Weibull or exponential) used for the transition and the number of parameters needed.
- The population priors.
- The number of state of deterioration and therefore the number of transitions between states.
- The number of asset groups (or types) and the factors (e.g. material used in construction) used to define membership of a group.
- The similarity of each group to the population as a whole.
- The remaining factors that adjust the transition distribution, possibly varying by group, and the weighting used to aggregate the effect of these factors.
- The types of data available.

Although these factors are uncertain they are not part of the probabilistic reasoning. Instead, these are the decisions made by domain experts to apply the generic models to a specific situation. The next two subsections describe how this is done: the first covers ‘modelling assumption’ and the second asset classification. Together, as shown in Figure 5, these processes turn the abstract probabilistic model into a model that can be run.

3.3 Customising the Abstract PRM with Modelling Assumptions

This aspect of the customisation covers four issues: a) the choice of transition distribution; b) the number of deterioration states; c) prior distributions and d) available inspection data.

3.3.1 Transition distribution and parameters

Different distributions can be used to estimate transition times, based on their goodness of fit. The goodness of fit of the distribution is usually done by visual observation and hypothesis test, such as coefficient of determination ($R^2$) and Anderson Darling (AD) test (Mendenhall et al., 2012). A range of study has been developed to find the best fit distribution of an asset state sojourn times. For example, the exponential distribution has been used for railway track (Guler et al., 2011) and the Weibull distribution for bridges (Sobanjo, 2011). The number of parameter in the distribution’s survival function fixes the number of instances of the Parameter class for each Transition. An example is showed in the left side of Figure 6: there are two instances of class Parameter if the guideline shows a Weibull distribution is normally adopted in practice.

3.3.2 Deterioration states used for grading

Each asset is usually rated with a state representing its functionality. For example, a 4 point grading system is used in Sobanjo (2011). Grading systems are normally adopted from industry standards and are often used to identify and priorities maintenance actions. They vary for different infrastructure type, countries, and sometimes, inspection agencies. An $n$-states grading system results in $n-1$ transitions represented in the instantiation of class Transition. An example is...
showed in Figure 6’s right side: there are two instances of class Transition since it is rated by a three-state grading system.

3.3.3 Asset deterioration characteristics: prior
Classical statistical methods, such as maximum likelihood estimators or least square method can be used to estimate priors if the data are sufficient. An alternative source is the expertise from experienced engineers (Welte and Eggen, 2008) from whom a prior range can be elicited. In addition, each group of assets is also characterised by its degree of similarity to the overall population. These group parameters are modelled using a truncated error distribution, with a mean inherited from the value of the learnt hyper-parameters, and the elicited degree of similarity defining the variance.

3.3.4 Inspection data: inferring state sojourn time
Continuous monitoring can provide exact transition times but periodic inspection is more common. In periodic inspection, the state of asset is only known at the inspection times. Therefore, the time an asset stayed in a state before deteriorating to another state is only constrained by the inspection result. Fortunately, this type of censored data can be modelled by different variants of the Data class. For example, periodic inspection requires three types: left, interval and right censored representing the state transition happening before, between and after the inspection respectively.

3.4 Asset Classification
Asset deterioration rate may be influenced by many contributing factors, such as age, loading and environment (Fu and Devaraj, 2008, Wellalage et al., 2014). By classify assets into groups with the same factor levels, we can expect groups to have similar patterns of deterioration (Veshosky et al., 1994). Our framework provides two ways to adjust deterioration based on such factors:

1. **Grouping**: some factors are chosen to define groups of assets so that historical data can be pooled and used to learn assets deterioration parameters. The number of groups fixed the number of instances of the Group class in the abstract probabilistic model.

2. **Adjusting**: other factors are used to adjust the deterioration rates learnt from data. A range of studies has been developed to identify these impact factors, for example, closeness to the coast, galvanic response level and structure type are identified in Yianni et al. (2016) as the key contributing factors for railway bridge deterioration. The number of these factors fixes the number of instances of the class Factor. The importance of each factor in the aggregated effect is modelled by a weight and this is used to shift effect of the learnt parameters.

3.5 Model Instantiation
The final stage of the process shown in Figure 5 is the instantiation of the BN. This set is provided by the framework as the domain expert has expressed all the information needed in the steps described in sections 3.3 and 3.4.

As Figure 3 shows, the customisation process also defines the schema of the database that holds both configuration parameters and the records of assets and inspections results. The class of each asset is defined either directly or inferred from the values of factors held in the database.

Suppose that investigated asset is \(x\), where \(x\) belongs to group \(g\), which is similar to group \(h\). We suppose that some time has elapsed since \(x\) last inspected and we wish to estimate its current state. The following steps are involved to instantiate the BN:

1. Creating transition variables for each deterioration stage of asset classes \(g\) and \(h\).
2. Creating hyper-parameters for each transition of group and at the population level. The group parameters approximate the population ones, using the similarity degree defined for the group.
3. Create the variables for the state of \(x\) and its adjustment by the factors (the ones that do not define the grouping). The values of the factor variables are extracted from the database and used as observations in the probabilistic calculation, but note that the model still operates if any of the values are missing (provided that priors have been provided).
4. Create variables for all the inspection data – taken from the database – for both groups \(g\) and \(h\). Starting from the current state, the inspection reports show either that one transition has occurred since the previous inspection or that no transition has occurred. By selecting the appropriate variant of data object both of these observations can constrain the transition time. It is even possible for more than one transition to have occurred between inspections.

For illustration, we assumed that it was known that group \(h\) needed to be include alongside data for assets in group \(g\), the group to which \(x\) belongs. Two steps are involved in automating this. Firstly, we need to determine whether the number of observations in group \(g\) is sufficient to give a good estimate of the deterioration rate. We could evaluate this either from the variance of the learnt parameters or from a threshold value on the absolute number. The second step is to find the group that is closest in characteristics to \(g\),
perhaps by the proportion of the values of the factors that define the groups shared between the two groups.

3.6 Inference and Refinement

Inference of the ground BN is performed automatically in the model-based machine learning framework. As suggested previously, probabilistic programming languages with an extensive list of inference algorithms can be adopted. For example, when dealing with hybrid Bayesian network that contains both discrete and continuous variables, Gibbs sampling can be used.

The query – to predict the unknown state of asset x – has to be expressed in domain terms and translated to a query on a BN variable. The result is a probability distribution over the possible states and further ‘decision rules’ or guidelines are needed to determine an action. For example, a small probability of the worst state of deterioration may determine that an urgent inspection is required.

Model evaluation can be made by comparison of different variants of the model, with the metrics such as predictive accuracy or computational speed. Further refinement of data sources (e.g. from other source domain), expert knowledge (e.g. different experts or different types of expertise), and variations in the models (e.g. different groupings of assets) are possible. This process repeats until a level of acceptable performance is accepted or it exhausts all the resource. As a future study, with the success of automatic inference software (Bishop, 2013), the refinement process can be made automatically with a defined threshold in a model-based machine learning framework.

4 REPRESENTING ASSET DETERIORATION MODEL WITH DIFFERERNT STRUCTURE

Customised instantiation of this framework for practical uses have been developed, for example, general maintenance problems in Zhang and Marsh (2016), and rail bridge deterioration in Zhang and Marsh (2018). Since the focus of this paper is to show how the model-based asset deterioration framework to deal with different modelling assumptions that may happen in practice, we present a series of instantiation variants of a basic deterioration model, which is sufficient to convey the idea.

4.1 The Simple Deterioration Model

The basic deterioration model is showed in Figure 1, and Table 1 shows its customisation setting. Notes that since the focus of this section is the model structure, we only show the setting that have influence on the shape of the model structure.

From the practice guideline, we decide there is only one transition (because it only has a binary state: working or failed), one group (representing the entire group has only one subgroup, the prior of the hyper-parameter becomes the prior of parameter), and one parameter (because its distribution is exponential). The expression of transition follows the distribution parameter in Parameter class, while the parameter’s prior is given by experts, here is a uniform distribution between 0 and 0.05. Four sojourn time data are inferred from the inspection data, and three of them have censored.

Based on this model, a range of variants can be extended to represent different assets, whose underlying modelling assumptions vary.

4.2 Variant 1: Deterioration that follows a two-parameter distribution

Extended from the basic model, Figure 7 shows a deterioration model of asset that follows a two-parameter distribution. Studies have found that transition probabilities between states of some assets are better fitted with distributions with two or even more parameters. For example, the two-parameter Weibull distributions for bridge deterioration are suggested in Ng and Moses (1998) and Sobanjo (2011).

This is achieved by the one-to-many relationship encoded in the relational database. Each instance of Transition class is linked to two identical instances of Parameter class.

4.3 Variant 2: Deterioration Under Multi-States

Asset may degrade with several stages representing the decrease of its functionality. For safety and relia-

Table 1. The setting of the model in Figure 1.

<table>
<thead>
<tr>
<th>Customisation Exponential (λ)</th>
<th>Grading system</th>
<th>Prior</th>
<th>Sojourn time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential</td>
<td>Binary states</td>
<td>λ ~ Uniform (0, 0.05)</td>
<td>15&lt;sojourn&lt;35</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>12&lt;sojourn&lt;24</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>12&lt;sojourn&lt;36</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>sojourn = 24</td>
</tr>
</tbody>
</table>

Figure 7. Deterioration follows a two-parameter distribution.
bility reasons, they can be rated from perfectly working to completely failure accompanied with several intermediate states. Extended from the last two subsections, Figure 8 shows two deterioration models for multiple states asset.

Degradation of asset with multi-states is modelled in the form of a Markov chain (Figure 8 (a)) by a sequence of states (represented by the transition nodes) representing the condition of an asset over time. Markov chain deterioration models are widely accepted in modelling most asset’s life-cycle performance (Frangopol et al., 2004), but they also bear with the assumption that the transition probabilities between states follow the same stationary transition rate, which do not change over time. This property implies the sojourn time follows an exponential distribution regardless how long it has been in the current state (Ng and Moses, 1998). This is performed by the one-to-many relationship in the relational database: one instance of Asset class with two instances of transition class.

This restriction can be relaxed by semi-Markov model (Figure 8 (b)), which allows the modelling of transition probabilities to follow non-stationary distributions depending on current state and its next state. This extension enables the modelling of multi-state deterioration that follows multi-parameters based distributions. This modelling is performed by the many-to-many relationship: one instance of Asset class with two instances of transition class, and each instance of transition class links to two instances of parameter class.

4.4 Variant 3: Learning from Similar Assets

Assets classified into different groups may share similar deterioration rate, which gives a potential to learn from others. Two types of learnings from similar assets are presented:

1) Figure 9 shows an example of pooling all the available data to learn a universal distribution in the domain, and distinguish a specific asset by defining the influence of aggregated external factors on deterioration rates. A suggestion use of this form is in the situation when the entire population has little data. This is achieved by the instantiation of Factor class.

2) Figure 10 shows an example of pooling available data within its associated group to learn their own distribution but governed by their shared hyper-parameter introduced by hierarchical BN. This hyper-parameter helps the transfer learning of the weekly learnt group (typically target group with little data) from strongly learnt group (source groups with lots of data). A suggestion use of this form is in the situation when the groups are highly correlated. This is achieved by the introduction of the class hierarchy by adopting layer 3 data sources.

5 CONCLUSION AND FUTURE STUDY

We have argued the need to provide a generalised asset deterioration framework encoded by probabilistic relational models, which can be adapted to model assets with different modelling assumptions. The emerging field of model-based approach gives us a suitable formalism for separating specifications from analysis techniques, and we have applied this to asset deterioration.
We also used several variants of the deterioration model to demonstrate how it can be adapted to a variety of applications with different modelling assumptions. In the future, we hope to extend the amounts of models in the model library, and extend the applications to more safety and reliability related problems.

ACKNOWLEDGEMENT

This work is supported by European Research Council (Funding code: ERC-2013-AdG339182-BAYES_KNOWLEDGE), and Agena Ltd for software support. H.Z. is supported by China Scholarship Council (CSC)/Queen Mary Joint PhD scholarships.

REFERENCE


