

Data-Driven Revision of Decision Models

Martin Žnidaršič

Jožef Stefan Institute, Slovenia

Marko Bohanec

Jožef Stefan Institute, Slovenia

Blaž Zupan

University of Ljubljana, Slovenia, and Baylor College of Medicine, USA

D

INTRODUCTION

Computer models are representations of problem environment that facilitate analysis with high computing power and representation capabilities. They can be either inferred from the data using data mining techniques or designed manually by experts according to their knowledge and experience. When models represent environments that change over time, they must be properly updated or periodically rebuilt to remain useful. The latter is required when changes in the modelled environment are substantial. When changes are slight, models can be merely adapted by revision.

Model revision is a process that gathers knowledge about changes in the modelled environment and updates the model accordingly. When performed manually, this process is demanding, expensive and time consuming. However, it can be automated to some extent if current data about the modelled phenomena is available. Data-based revision is a procedure of changing the model so as to better comply with new empirical data, but which at the same time keeps as much of the original contents as possible. In the following we describe the model revision principles in general and then focus on a solution for a specific type of models, the qualitative multi-attribute decision models as used in DEX methodology.

BACKGROUND

The task of data-driven revision is to adapt the initial model to accommodate for the new data, while at the same time making use of the background knowledge, which was used in the construction of the initial model. Revision is to be applied only when the changes of the

modelled concepts are not substantial, that is, if we deal with concept drift (Tsymbal, 2004). If the changes of the modeled system are substantial, it is usually better to construct the model from scratch.

Depending on the field of research, procedures of this kind are most often referred to as knowledge refinement or theory revision. Most of research in this field is done on propositional rule bases (Ginsberg, 1989; Mahoney & Mooney, 1994; Yang, Parekh, Honavar & Dobbs, 1999; Carbonara & Sleeman, 1999) and Bayesian networks (Buntine, 1990; Ramachandran & Mooney, 1998), but many principles of these methods are shared with those of revision procedures for other knowledge representations, such as case-based reasoning systems (Kelbassa, 2003).

Multi-criteria decision models (MCDM) are models used in decision analysis (Clemen, 1996). Data-driven revision can be a valuable tool for the ones that are used for longer periods of time. In our previous work, we have developed revision methods (Žnidaršič & Bohanec, 2005; Žnidaršič, Bohanec & Zupan, 2006) for two types of MCDM models of DEX methodology (Bohanec & Rajkovič, 1990; Bohanec, 2003). An input to MCDM models is criteria-based description of alternatives, where a model represents a utility function to evaluate the given set of criteria values. MCDM models are used for evaluation and analysis of decision alternatives. In contrast to traditional numerical MCDM (Saaty, 1980; Keeney & Raiffa, 1993; Triantaphyllou, 2000), the models of DEX methodology are qualitative and have utility functions in form of if-then rules. The concepts in these models are structured hierarchically and their values are defined according to the values of their immediate descendants in the hierarchy (see Figure 1). This dependency is specified with qualitative rule-based utility functions, which can be defined

as crisp or probabilistic and are usually represented in tabular form (see Table 1). The concepts at the bottom of hierarchy serve as inputs, represent the criteria-based description of alternatives and must be provided by the user.

Models of DEX methodology are usually constructed manually in collaboration of decision analysts and problem domain experts. As an alternative, a method called HINT (Bohanec & Zupan, 2004) was proposed that can infer DEX-like models from data. HINT often requires a large quantity of data and its discovery process may benefit from an active involvement of an expert. The task of revision is simpler and can be achieved completely autonomously with very limited amount of new evidence.

MAIN FOCUS

Revision Goals

For any kind of knowledge representation or model (M), the goal of revision methods (r) is to adapt the model with respect to new evidence from the changed environment. In the case of data-driven revision, evidence is in the form of a set of data items ($D = \{d_1, d_2, \dots, d_n\}$). The success of adaptation may be demonstrated

through the increase of some selected measure (m) that assesses the performance of the model given a set of test data. While standard measures of this kind are, for instance, classification accuracy and mean squared error (Hand, Mannila & Smyth, 2001), any other measure that fits the modelling methodology and problem may be used.

The quality of the model after the revision should thus increase:

$$m(r(M, D), D) \geq m(M, D),$$

where it aims to maximize: $m(r(M, D), D)$.

However, revision is a process that tries to preserve the initial background knowledge, rather than subject it entirely to the new data. The maximization from the latter equation must be therefore limited by the type and degree of changes that the revision method is allowed to make. If there exists background data that was used in the initial model construction (D_b) or we are able to reproduce it, we can also limit the revision by fully considering the prior data:

$$m(r(M, D), D_b) \geq m(M, D_b),$$

or at least by minimizing the difference: $m(M, D_b) - m(r(M, D), D_b)$.

Figure 1. A simple hierarchy of concepts of a decision model for car purchase

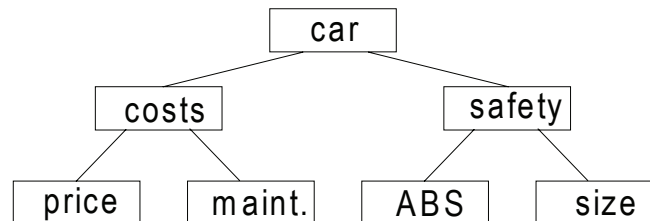


Table 1. Utility function of the concept safety

ABS	size	safety
no	small	<v.good:0.0, good:0.0, accep.:0.1, bad:0.9>
no	medium	<v.good:0.0, good:0.1, accep.:0.7, bad:0.2>
no	big	<v.good:0.1, good:0.8, accep.:0.1, bad:0.0>
yes	small	<v.good:0.0, good:0.0, accep.:0.3, bad:0.7>
yes	medium	<v.good:0.1, good:0.6, accep.:0.3, bad:0.0>
yes	big	<v.good:0.8, good:0.2, accep.:0.0, bad:0.0>

5 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/data-driven-revision-decision-models/10885

Related Content

Flexible Mining of Association Rules

Hong Shen (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition* (pp. 890-894). www.irma-international.org/chapter/flexible-mining-association-rules/10925

Storage Systems for Data Warehousing

Alexander Thomasian (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition* (pp. 1859-1864). www.irma-international.org/chapter/storage-systems-data-warehousing/11072

Data Preparation for Data Mining

Magdi Kamel (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition* (pp. 538-543). www.irma-international.org/chapter/data-preparation-data-mining/10872

Frequent Sets Mining in Data Stream Environments

Xuan Hong Dang, Wee-Keong Ng, Kok-Leong Ong and Vincent Lee (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition* (pp. 901-906). www.irma-international.org/chapter/frequent-sets-mining-data-stream/10927

On Interactive Data Mining

Yan Zhao (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition* (pp. 1085-1090). www.irma-international.org/chapter/interactive-data-mining/10956