



Enhancing ontology-based antipattern detection using Bayesian networks

Dimitrios Settas^{a,*}, Antonio Cerone^a, Stefan Fenz^b

^aUnited Nations University, International Institute for Software Technology, Macau, SAR, China

^bInstitute of Software Technology and Interactive Systems, Vienna University of Technology and SBA Research, Vienna, Austria

ARTICLE INFO

Keywords:

Bayesian networks
Ontology
Antipatterns
Antipattern detection

ABSTRACT

Antipatterns provide information on commonly occurring solutions to problems that generate negative consequences. The antipattern ontology has been recently proposed as a knowledge base for SPARSE, an intelligent system that can detect the antipatterns that exist in a software project. However, apart from the plethora of antipatterns that are inherently informal and imprecise, the information used in the antipattern ontology itself is many times imprecise or vaguely defined. For example, the certainty in which a cause, symptom or consequence of an antipattern exists in a software project. Taking into account probabilistic information would yield more realistic, intelligent and effective ontology-based applications to support the technology of antipatterns. However, ontologies are not capable of representing uncertainty and the effective detection of antipatterns taking into account the uncertainty that exists in software project antipatterns still remains an open issue. Bayesian Networks (BNs) have been previously used in order to measure, illustrate and handle antipattern uncertainty in mathematical terms. In this paper, we explore the ways in which the antipattern ontology can be enhanced using Bayesian networks in order to reinforce the existing ontology-based detection process. This approach allows software developers to quantify the existence of an antipattern using Bayesian networks, based on probabilistic knowledge contained in the antipattern ontology regarding relationships of antipatterns through their causes, symptoms and consequences. The framework is exemplified using a Bayesian network model of 13 antipattern attributes, which is constructed using BNTab, a plug-in developed for the Protege ontology editor that generates BNs based on ontological information.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

Applying a previously successful solution to a similar problem is one of the established ways in which humans solve problems and is the main concept behind design patterns. After the emergence of design patterns, the pattern community slowly developed problem–solution pairs in which the solution is not beneficial and leads to further difficulties and problems. Antipatterns are the next generation of design pattern research and are related with patterns in the sense that design patterns can evolve into antipatterns. An antipattern is a new form of pattern that has two solutions. The first is a problematic solution with negative consequences and the other is a refactored solution, which describes how to change the antipattern into a healthy solution (Brown, McCormick, & Thomas, 2000). The second solution is what makes antipatterns beneficial. The difference is in the context: An antipattern is a pattern with inappropriate context and is particularly useful in the case of knowledge representation, because it captures experience and provides information on commonly

occurring solutions to problems that generate negative consequences (Brown, Malveau, McCormick, & Mowbray, 1998). The process that is followed by a pattern to change its solution into a better one is called refactoring. This solution has an improved structure that provides more benefits than the original solution and refactors the system toward minimized consequences. There exist different categories of antipatterns. For the purposes of this paper, software project management antipatterns are used in order to exemplify the proposed approach.

In previous work, Bayesian Networks (BNs) (Settas, Bibi, Sftos, Stamelos, & Gerogiannis, 2006) and the formalism of ontology (Settas & Stamelos, 2007; Settas, Meditskos, Stamelos, & Bassiliades, 2011) have been used separately in order to produce statistical and extensible ontological models of antipatterns. Bayesian networks provided a framework for project managers, who would like to model the cause–effect relationships that underlie an antipattern, taking into account the inherent uncertainty of a software project. An antipattern BN model provides the precise mathematical model of a project management antipattern and can be used to measure and handle uncertainty in mathematical terms. By applying the formalism of ontology (Settas & Stamelos, 2007; Settas et al., 2011), a common lexicon of terms that can be communicated across people and software tools was defined. This provided the

* Corresponding author.

E-mail addresses: dsettas@csd.auth.gr (D. Settas), antonio@iist.unu.edu (A. Cerone), stefan.fenz@tuwien.ac.at (S. Fenz).

basis for the application of further methodology in order to address similar antipattern ontologies and implement SPARSE (Settas et al., 2011), an ontology-based intelligent system that can detect antipatterns based on the symptoms that appear during a software project. This system has recently been proposed for teaching and learning using antipatterns (Settas & Cerone, 2011).

In this paper, we intertwine the formalisms of Bayesian Networks (BNs) and Ontologies. Ontology is a methodology that has been widely used for the development of expert systems (Liao, 2005). Bayesian networks are also widely used in expert systems for prognosis and prediction purposes (Ferreiroa, Arnaiza, Sierrab, & Irigoinb, 2012). The research goal in this paper is to further develop the existing antipattern ontology knowledge base by using Bayesian network analysis. Uncertainty concerns every aspect of ontologies (Taniar & Rahayu, 2006; Pan, Ding, Yu, & Peng, 2005) and it is one of the most important criteria that need to be taken into account in the selection of an appropriate knowledge representation (Richardson & Domingos, 2003). The task of capturing incomplete or uncertain antipattern knowledge is essential but is often not explored based on assumptions on the certainty and accuracy of software project data. For example, the certainty regarding the existence of the relationship between a specific cause and symptom of an antipattern. Support for uncertainty is essential for intelligent systems that support antipatterns, such as SPARSE because the creation of new antipatterns using the antipattern OWL ontology often relies on information from past project which comes from experience, memory and intuition. Antipattern ontology contributors might often rely on their own expert judgement to define how antipatterns might be linked, which has a clear impact on the effectiveness of the antipattern detection process. By incorporating BNs in the antipattern OWL ontology, we can represent probabilistic information regarding antipatterns and their attributes. As a result, the detection process can be improved by providing users not only a set of antipatterns that exist in a software project based on the symptoms they selected, but also to quantify, measure and visualize the uncertainty that lies behind antipattern attributes and relationships.

At the moment there exist a variety of ontology editors offering different functionality. Protege is one of the most widely used open source ontology editors and was chosen for the implementation of SPARSE and the antipattern OWL ontology (Settas et al., 2011). Further development of the ontology in previous work tackled the challenge of providing a web-based collaborative environment by using WebProtege¹ (Tudorache, Vendetti, & Noy, 2008), which is a web-client for Collaborative Protege. There is a variety of frameworks and tools under development in order to support the study of uncertainty in ontologies. BayesOWL framework (Ding, Peng, & Pan, 2004; Pan et al., 2005) can translate an OWL ontology into a Bayesian Belief Network structure, but it has still not been fully developed. Other frameworks currently under development such as POWL (Hung, Szeto, & Deng, 2009) and PR-OWL (Carvalho, Laskey, & Costa, 2010) can extend OWL vocabulary for representing uncertainty in different expressivities. These frameworks are being studied by the Uncertainty Reasoning for the World Wide Web Incubator Group (URW3-XG) (Laskey & Laskey, 2008) but at the moment they lack compatibility with OWL. In this paper, the issue of quantifying uncertainty in the antipattern ontology is addressed by using the BNTab Protege plugin. BNTab is the only plug-in that is currently available for Protege that can generate Bayesian networks based on ontologies. The BNTab² plug-in enables users to efficiently generate Bayesian networks based on existing ontologies.

The main contribution of this paper is the application of Bayesian networks on the antipattern OWL ontology. This approach

incorporates probabilistic information in the ontology and allows the semi-automatic generation of BNs using BNTab. In that way, BNTab can illustrate the uncertainty that exists in antipatterns and antipattern attributes and enhance the antipattern detection process. Ultimately, the BNTab plugin will be used within the web-based collaborative version of the antipattern ontology (Settas, Stamelos, & Bassiliades, 2011) and antipattern contributors will be able to use the approach proposed in this paper simultaneously through the Web. This will ensure that all antipattern ontology users are editing the most up-to-date version of the ontology. SPARSE can benefit from this approach by including associated BNs with the resulting detected antipatterns in order to visualize the uncertainty of the data used for the detected antipatterns.

This paper is divided in six sections, which are organized as follows: Section 2 describes the background, the related work and the literature review used in our research. Section 3 introduces the reader to Bayesian networks and the BN analysis used in the remaining sections. Section 4 describes the antipattern ontology and presents a short description of the rules and reasoning used in SPARSE in order to detect antipatterns. Section 5 presents the process of extending the OWL ontology in order to describe probabilistic information. Furthermore, the application of BNTab in the antipattern ontology is described and the approach is exemplified using an example of a software project management antipattern. Finally, in Section 6, the findings are summarized, future work is proposed and conclusions are drawn.

2. Background and related work

2.1. Background

Bayesian networks and ontologies are different knowledge representation formalisms. In this paper these formalisms are intertwined, therefore it is important to understand how they can supplement each other at a syntactical but also inferential level. At a knowledge representation level, the antipattern ontology can determine the categories of things that exist in an antipattern and set the ontological commitments of the project manager, system architect or application designer (Davis & Szolovits, 1993). Both Ontologies and BNs offer a medium for efficient computation because they do not only represent knowledge, but also encode knowledge in a form that can be processed efficiently by software. Furthermore both BNs and ontologies offer a medium for human expression that can be used by project managers in order to have a common lexicon of terms with a solid mathematical foundation. The antipattern ontology encodes tacit software project management knowledge into a computer understandable form and allows the sharing and reuse of this knowledge by software tools, such as SPARSE which is a symptom-based antipattern retrieval knowledge-based system using semantic Web technologies.

Both BNs and ontologies have the ability to carry out reasoning. In the case of BNs, reasoning is carried out under uncertainty and by illustrating the uncertainty that surrounds antipatterns. The Bayesian capacity to draw strong inferences from sparse data cannot be ignored. However, it is the formalism of ontology that provided a sound basis for the further development of methods and software tools using ontology. By incorporating a BN model of an antipattern inside the ontology, managers can still take advantage of the mathematical analysis of BNs while using the formalism of ontology. The antipattern ontology served as the knowledge base for the development of the intelligent system (Settas et al., 2011) that can detect the antipatterns that exist in a software project.

A knowledge representation language has a syntactic and an inferential aspect. The syntactic aspect is notational and refers to

¹ <http://protegewiki.stanford.edu/wiki/WebProtege>

² <http://stefan.fenz.at/ontology-based-generation-of-bayesian-networks/>

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات