Explanations in Bayesian Networks using Provenance through Case-based Reasoning

Anders Kofod-Petersen, Helge Langseth, and Agnar Aamodt

Department of Computer and Information Science,
Norwegian University of Science and Technology,
7491 Trondheim, Norway
{anderpe|helgel|agnar}@idi.ntnu.no

Abstract. Bayesian Networks are useful for solving a wide range of problems in many domains. Yet, they are exposed to one important challenge when structural and parametrical changes occur. As Bayesian networks lack memory regarding changes over time, there is currently no good way of maintaining a history of changes and their provenance. Thus, any variance in the network's problem solving behaviour will not be explainable to a user. Within the context of systems that integrate case-based reasoning and Bayesian networks, we suggest to add a case-based reasoning functionality that will retain changes and their provenance, as well as approaches to explain any unexpected problem solving behaviour.

1 Introduction

Explanations have been identified as one of the most important properties of intelligent systems (see e.g. [1,2]). One important source of knowledge for generating useful explanations is the information related to changes in used knowledge sources and changes in reasoning processes.

Although provenance-related information has been included in some case-based reasoning (CBR) systems since the early days of the field, and often referred to as book-keeping information, justifications, or meta-information: a systematic treatment of provenance and its role was introduced by Leake and Whitehead [3]. They identified two main roles of provenance in case-based reasoning; to compensate for delayed feedback in the REVISE phase of the case-based reasoning cycle [4], and to improve case base maintenance. The role suggested here is an additional one, namely to produce explanations for unexpected results that stem from changes in parts of the domain model. In a philosophical treatment of provenance and explanation [5] the focus is on "how" and "why" type explanations relate to provenance, and in a treatment of explanation and provenance within a Description Logics framework. An emphasis is on how to track "how-provenance" in a system, referring to how sources of provenance contribute to the observed data.

In the work reported here the domain model could in principle be a case-specific domain model as well as one containing generalised knowledge, but in the present paper we focus solely on changes in the general domain knowledge, and more specifically in the Bayesian network (BN) component of a system.

Bayesian networks [6,7] constitute a modelling framework for uncertain knowledge. With their formal grounding in probability theory and statistical inference, Bayesian networks have become among the most used frameworks for reasoning under uncertainty, and has found applications in areas as diverse as genetics, failure detection, recommender systems, and speech recognition. One important argument for using Bayesian networks is its ability to adapt both its structure [8] and its parameters [9] as new training data is presented to the system; properties which are crucial in domains that are partly unknown or changing.

The nature of Bayesian networks also allows for some explanations to be given regarding the reasoning process (see Lacave and Díez [10] for a review). However, there is one important aspect where Bayesian networks cannot explain changes. That is when the perceived behaviour of the system has changed via learning (via adaptation of the Bayesian network, or – in a hybrid system – by adding cases to the case base). In other words, a Bayesian network might experience changes in structure, e.g. new observable parameters can appear, or changes in parameters, e.g. a change in the probability distribution in the probability function in a node. As Bayesian networks traditionally retains no history of the provenance of such changes made during learning, any variation in results coming from such changes are inexplicable.

In previous work [11], we have been looking into the benefits of combining Bayesian networks and case-based reasoning into a hybrid system, with the goal to utilise the Bayesian networks either as a reasoning engine in its own right, or to help with the case matching in the integrated system. This is an endeavour we have taken up again recently, and although in this paper we will exemplify the reasoning processes using only a Bayesian network, we do still have the total hybrid system in mind.

The work presented here suggests the use of case-based reasoning to address the lack of ability to explain changes in the perceived behaviour of systems including Bayesian components. In brief, we propose to construct cases based on changes in the Bayesian network and use this in combination with a model of the user to construct explanations. Sørmo et al. [12] identified five explanation goals: Explain How the System Reached the Answer (Transparency), Explain Why the Answer is a Good Answer (Justification), Explain Why a Question Asked is Relevant (Relevance), Clarify the Meaning of Concepts (Conceptualization), and Teach the User About the Domain (Learning). Our focus here is on the transparency and justification goals.

The rest of this paper is organised as follows: Section 2 introduces a motivational example for augmenting Bayesian networks with a case-based reasoning provenance part; Section 3 describes how case-based reasoning can be applied and how the cases can be constructed; The paper ends with a summary and an outlook on future work.

2 Motivational Example

To exemplify the need for provenance of changes in a Bayesian network we will use a very simple example. A simple Bayesian network, as depicted in Fig. 1, can be used to tell a user whether to go to the beach or not. Real world application would probably be much more complicated and even make recommendation without being explicitly asked to do so. However, for pedagogical reasons we will stick with simplicity.

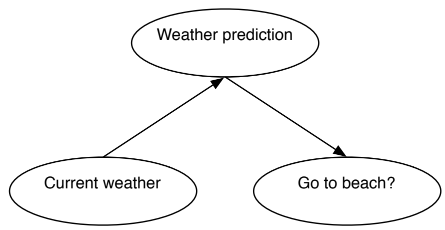

Fig. 1. Bayesian network

The network can observe the current weather and if it is raining or cloudy will tell the user to stay at home, and if it is sunny it will tell the user to go to the beach. The column marked 'Historical recommendation' in Table 1 shows the mapping between weather conditions and expected recommendations. This explicit mapping is naturally not visible to the user. However, we can assume that a user who uses any system on a regular basis will develop some expectation as to how they behave.

Table 1. Condition and response

Condition	**Historical recommendation**	**New recommendation**
Rain	Stay at home	Stay at home
Cloudy	Stay at home	Go to beach / stay at home
Sunny	Go to beach	Go to beach

Let us assume that the system now gets equipped with a barometer, which to some degree can be used to predict the weather. The new Bayesian network, including the barometer is depicted in Fig. 2. Using this new device the network is able to predict that cloudy weather in some situations will improve and become sunny. Thus, the system will sometimes suggest the user to go to the beach even when it is cloudy, for example when the barometer shows a trend towards higher pressure. This new observed behaviour is shown in the column marked 'New recommendation' in Table 1.

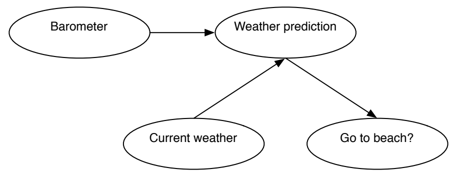

Fig. 2. Modified Bayesian network

The first time the system makes an *unexpected* prediction, i.e. suggests going to the beach even when it is cloudy, the user might like to see an explanation of the apparent inconsistency. A Bayesian network can supply an explanation regarding the relation between the observed variables (the current weather and the barometer reading), the weather forecast, and the query variable. However, as aforementioned no explanation is readily available regarding the fact that there has been a structural change. The types of explanation available from the Bayesian net are clearly not sufficient as they do not include the real provenance of the change in the reasoning process. Since the system behaves in an unexpected manner, the user is likely to be interested in a justification of the change in the expected outcome.

3 Case-based Reasoning as an Explanation Mechanism

The work presented here suggests to augment Bayesian networks with a case-based reasoning component to explain changes in behaviour. This is achieved by retaining the provenance of either procedural or structural changes in the problem-solving network.

When a request for an explanation is received, the case-based reasoning system will retrieve a case containing a relevant part of the previous network that solved the problem (in the example predicting whether or not to go to the beach), and the current network. New cases are constructed whenever a user asks a question to the system. In this example, when a request for a recommendation for visiting the beach is made. If the answer to the question and the input parameters are sufficiently similar to an existing case, no new case is retained. The request for an explanation indicates an unexpected solution or insufficient confidence in the suggested solution, for which a transparency and/or justification explanation is called for. Figure 3 depicts an example case.

In the problem description part of a case, the findings are: a user model, which can be used to influence the explanation goal to be solved and the way to present it; the current network; and the last know accepted network (marked in

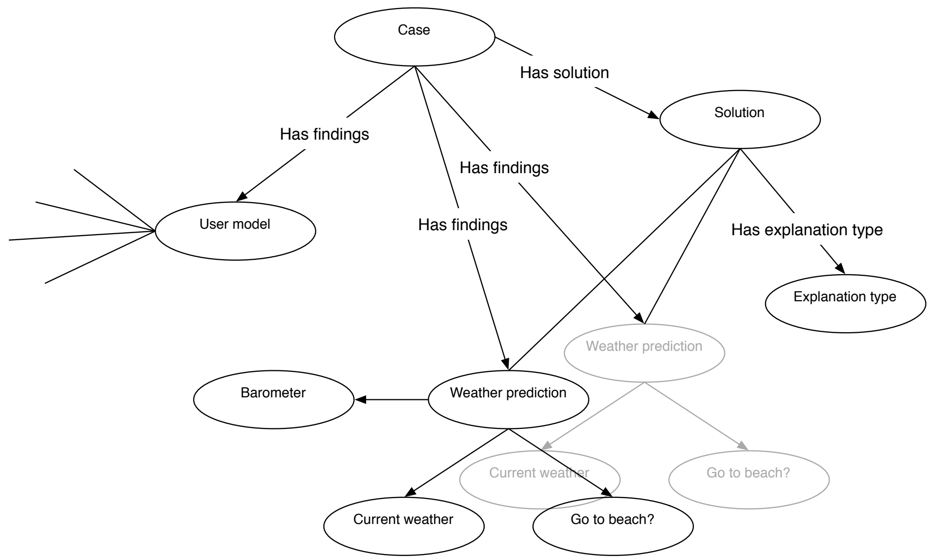

Fig. 3. Case structure

light grey). The solution to the case contains the explanation goal and the two Bayesian networks. The latter can be used as a knowledge container [13] where relevant knowledge to construct the explanation can be acquired.

The user model contains information about the user in question. Typically it would contain some user identifier, and importantly information about the level of skills of the user. This information is important when deciding the type of explanation to provide. As an example, if the user is a highly skilled expert the *transparency* explanation is likely to be the preferred one, where as for novices a *justification* is more likely to be preferred (for a relationship between user competencies and transparency versus justification, see e.g. [14]).

In the case of the user requesting an explanation, as an example in the form of: "The last time it was cloudy, you told me to stay at home. However, now you are telling me to go to the beach. Why?" The current case is matched against the last case recommendation expected by the user. The current case contains the current Bayesian network, which is then matched against the network in the retrieved case. The difference between the two cases from the basis for the explanation.

To continue the example from Section 2 and augment it with the case-based reasoning provenance system, we can now supply an explanation as to why the answer differs from the expected one. Comparing the two networks in question (see Fig. 3) shows that the current network contains a new node, the barometer. This barometer affects the system's ability to predict the weather. Given that it is cloudy and the barometer says that there is a high pressure, it will presumably be sunny shortly. So the explanation offered by the system might be: "Since the last time that cloudy weather was observed, when I suggested to stay at home, I have received a barometer. The barometer tells me that even though it is cloudy, there is a high pressure. Thus, it is very likely that it will be sunny at the beach."

4 Summary and Further Work

We are currently investigating suitable approaches to implement the suggestions made in this paper. A suitable implementation in a domain closer to the real world must be carried out in order to test the validity of this Bayesian networks and case-based reasoning hybrid approach.

So far we have suggested how augmenting the Bayesian network with the case-based reasoning system allows the system to supply a *transparency* explanation, to e.g. an expert, and a *justification* explanation, to e.g. a novice.

It is likely that with a suitable structuring of the knowledge model, other types of explanation are possible, either directly from the Bayesian network or from the case-based reasoning provenance system. These types of explanations could include the complete list from [12]. Different types of explanations satisfies different goals [2]. Among these are, perhaps the least explored, namely *learning*. The learning type allows a user to learn about the domain in question. This might be a two-way street, where the user might be able to teach the system why a new solution is worse that an old one. Combined with the other types of explanations –where the user is the producer and the system the consumer– changes to the Bayesian network might be suggested to the system. However, this conversational perspective on explanations is in an early stage, and for far remains unexplored.

Finally, other types of relevant provenance information, such as the context of the problem-solving system and the rationale for changes in the problem-solver's recommendations, could be combined with the suggested approach to further improve users' trust and acceptability of reasoning systems.

References

1. Leake, D.B.: Evaluating Explanations: A Content Theory. Lawrence Erlbaum Associates (1992)
2. Roth-Berghofer, T.R., Cassens, J.: Mapping goals and kinds of explanations to the knowledge containers of case-based reasoning systems. In Muñoz-Avila, H., Ricci, F., eds.: Case Based Reasoning Research and Development – ICCBR 2005. Volume 3630 of LNAI., Chicago, Springer (2005) 451–464
3. Leake, D., Whitehead, M.: Case provenance: The value of remembering case sources. Case-Based Reasoning Research and Development (2007) 194–208
4. Aamodt, A., Plaza, E.: Case-based reasoning: Foundational issues, methodological variations, and system approaches. AI communications **7**(1) (1994) 39–59
5. Lockie, R.: Knowledge, Provenance and Psychological Explanation. Philosophy **79**(03) (2004) 421–433

6. Pearl, J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann (1988)
7. Jensen, F.V., Nielsen, T.D.: Bayesian Networks and Decision Graphs. 2nd edn. Springer (2007)
8. Friedman, N., Goldszmidt, M.: Sequential update of bayesian network structure. In: Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence, Morgan Kaufmann (1997) 165–174
9. Spiegelhalter, D.J., Lauritzen, S.L.: Sequential updating of conditional probabilities on directed graphical structures. Networks **20** (1990) 579–605
10. Lacave, C., Díez, F.: A review of explanation methods for bayesian networks. The Knowledge Engineering Review **17**(2) (2002) 107–127
11. Aamodt, A., Langseth, H.: Integrating Bayesian networks into knowledge intensive CBR. In: American Association for Artificial Intelligence, Case-based reasoning integrations; Papers from the AAAI workshop – Technical Report WS-98-15, Madison, WI., AAAI Press (1998) 1–6
12. Sørmo, F., Cassens, J., Aamodt, A.: Explanation in case-based reasoning – perspectives and goals. Artificial Intelligence Review **24**(2) (October 2005) 109–143
13. Richter, M.M.: The Knowledge Contained in Similarity Measures. Invited Talk at the First International Conference on Case-Based Reasoning, ICCBR'95, Sesimbra, Portugal (1995)
14. Kofod-Petersen, A., Cassens, J.: Explanations and context in ambient intelligent systems. In kokinov, B., Richardson, D.C., Roth-Berhofer, T.R., Vieu, L., eds.: Modeling and Using Context, proceedings of the 6th International and Interdisciplinary Conference (CONTEXT 2007). Volume 4635 of Lecture Notes in Artificial Intelligence., Roskilde, Denmark, Springer (August 2007) 303–316