

# CBR for Advice Giving in a Data-Intensive Environment

Agnar AAMODT  
*Volve AS, Trondheim, Norway*

**Abstract.** Reusing past experiences by reasoning from past cases poses particular problems when the input to case retrieval comes from large amounts of online data. Volve has developed a system in which data from oil well drilling logs are continuously monitored, interpreted, and used to check if previous incidents exist that may predict an unwanted event to happen unless preventive actions are taken. The system is currently being tested.

**Keywords.** case-based reasoning, decision support, online data analysis, oil-well drilling  
*Invited talk summary paper.*

## Introduction

Reusing past experience, i.e. being reminded of similar situations and making use of decision steps made earlier, has turned out to be an efficient way to handle new situations for human beings. Case-Based Reasoning (CBR) builds upon this principle in order to provide improved computer-assistance to people in their daily work. As CBR is becoming more utilized for experience capture and reuse in real-world industrial settings (e.g. [1]), methods for linking human experiences to the observed data related to those experiences are getting increased attention. Many industries, including the oil & gas industry, have access to large amounts of data and information, and advanced tools for displaying various types of information. As the amount of available data increases, the need for tools to extract, or filter out, the relevant information in a given situation increases correspondingly.

We are developing a system that will assist oil well personnel during drilling operations in improving the quality and efficiency of the drilling process. This development is part of a cooperation with the Norwegian oil company StatoilHydro, which has enhanced drilling performance as one of its prioritized areas.

The current version of the system helps to avoid “unwanted events”, i.e. events that lead to a slower drilling progression than expected. A particular focus of the system is to help with problems of “poor hole cleaning”, i.e. problems that lead to increased resistance for the drill string due to erroneous accumulation of material along the well wall.

Human experience from earlier poor hole cleaning incidents are gathered in a case base. The case base is linked to a model of general domain knowledge, in a manner somewhat similar to the method in the Creek system [2, 3]. In that system the assumption was that case features given as input to the system, i.e. the index terms characterizing a new situation, were entered as symbolic entities. The current system under development at Volve [4] is linked online to an ongoing drilling process, supervises the process by continuously collecting numerical and symbolic data from a large number of parameter readings, interprets these readings, retrieves one or more past cases that match the current state of the drilling process, and on that basis gives advice about how to proceed in order to avoid a possible unwanted event.

In the rest of this paper the steps from the reading of well data to the advices being given to a drilling professional are explained in more detail.

### 1. Data collection and interpretation

As part of their normal work oil well drilling engineers and other operational personnel both offshore and in support centers onshore have at their disposal a large set of sensor measurements and other drilling parameters. The main portion of these data are continuous data streams from the drilling operation. Tools for keeping track of data from these drilling logs help the personnel to perform graphical comparisons through time- or depth-indexed graphs. This kind of software can be quite powerful visualization tools, but they have limited capabilities in giving the user advices based on interpretation and analyzes of the data.

Volve’s software extends the screen information to ensure better decisions. One extension is by giving explicit high-level well status information based on the interpretation of the data. This is done by identifying and displaying particular “interpreted events” attached to the data logs, as the drilling process proceeds. These events are high level interpretations that characterize the status of the well. This is illustrated in the left part of Figure 1. The two leftmost columns are example data logs displaying parameter values (of block position and drill string torque, in this example) vs. time. The horizontal line below the log graphs indicate the current hole depth. The third column shows a sequence of the events interpreted by the software based on the streams of log data. The events below the horizontal line are events predicted by the system (stippled boxes refer to future entities). The middle and right parts of the figure illustrate the system modules responsible for the two main types of decision support functionalities:

- Prediction of unwanted events on a short time scale
- Predication of unwanted event on a longer time scale, with advice on preventive actions.

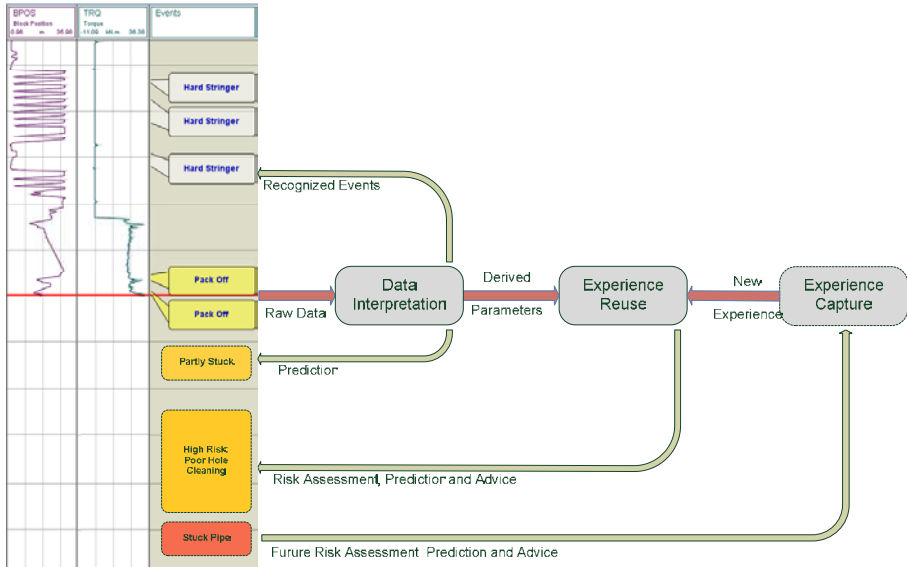


Figure 1. Excerpt from Volve system screen annotated with responsible method modules.

The Data Interpretation module ‘listens’ to the set of continuous data streams – typically 20-30 different logs. It has a set of methods, partly based on physical models, and partly on pattern analysis, that will suggest possible unwanted events directly from the data. This can only be done reliably in a short-time perspective – i.e. in the order of a few minutes. In the figure a Partly Stuck event has been predicted in this way. Typically, the user would be given an alarm or other kind of warning in a situation like this.

The Data Interpretation module also has a second role, in addition to short-term prediction: It interprets numerical log data into symbolic features such as qualitative parameter values, trends, interpreted activities, interesting events, etc. for the purpose of identifying useful features for the retrieval of relevant past cases. The cases, in turn, enable more long-term prediction of the well condition.

## **2. Case matching and reuse**

The case base captures the human experiences related to previous unwanted events as a set of past cases. The case base is the primary knowledge source for the Experience Reuse module (see Figure 1). Initially, the cases are described manually, based on the analysis of written drilling reports and other textual sources, manually interpreted data logs, and if needed combined with interviews of drilling personnel who were involved in the past incidents. As the system is being used, methods for automated learning of new cases, and automated case base maintenance policies, will take over. This is illustrated by the Experience Reuse module to the right in Figure 1.

The Experience Reuse module attempts to find a matching case (or set of matching cases), with a degree of match above a certain threshold. The case or cases are then used to make predictions of possible unwanted events on a longer time scale. These predictions are based on what happened in the past case(s) retrieved. The user may want to inspect more cases than immediately displayed by the system, by moving down the similarity threshold ladder. On the basis of an identified past case that is sufficiently similar to the current situation, actions are suggested to the operators that should be taken to avoid the predicted event.

A case in this system is a rich source of knowledge. It contains structured data in the form of parameter hierarchies, it contains numerical, symbolic, and textual data. The textual data is so far only human interpretable. A part of the case, corresponding to parameter values and context information available up to the point of time when the case was captured, is used as index features in the case matching. The rest of the case is used for advice giving after a satisfactory match has been found.

The similarity assessment method assumes that both numerical and symbolic parameters will be used in the matching. A particular challenge is posed by the fact that a realistic case matching process cannot be based on state descriptions in a single snapshot of time only, but needs to take sequence information into account as well.

After the system has predicted possible hazards, and the drilling operator has made the appropriate preventive actions and evaluated the result, the system should learn from its advice giving effort – whether the system’s advice turned out successful or not. Learning as part of the normal run-time operation is a characteristic of case-based reasoning, compared to other machine learning methods. Each problem solving experience, i.e. each prediction of a possible unwanted event made by the computer, is a powerful source for learning. The system’s learning from its own experience should at least ensure that a mistake once made is never repeated.

### 3. Status and Conclusion

Referring to figure 1, the Data Interpretation and Experience Reuse modules have been developed and are now undergoing testing. Testing is done partly by controlled tests on previous data logs, partly by coupling the system to ongoing drilling operations on platforms in the North Sea. The Experience Capture Module will be finalized at a later stage, when the other two modules are ready for field deployment and new cases will start to come into the system.

The tests made so far are very promising. Strong and active user support from StatoilHydro, combined with a highly iterative system development process, have ensured that requested adaptations in system functionality and user interface have been implemented as part of the normal development process.

As for CBR as a method for computerized advice giving and decision support in this type of environment, the need for methods that can deal with continuous streams of online data will be increasingly called for. Data abstraction methods need to be coupled with complex similarity assessment methods that go far beyond the traditional “feature vector“ methods. This problem has been recognized and addressed in the CBR community (e.g. [5]). In the work described here some contributions have also been made in this direction, although the main contribution lies in demonstrating the usefulness of our combined data interpretation and case-based reasoning approach in a highly data-intensive, and reactive industrial setting.

### Acknowledgements

The work reported here has been partly funded by StatoilHydro and the Research Council of Norway under contract no. 169463/S300, the RTDA project. Many people from StatoilHydro and Volve have contributed to the results presented, in particular the project officer at StatoilHydro, Erik Nyrnes, Volve’s CEO Jone Rasmussen, the Volve development leaders Pål Skalle and Frode Sørmo, and Tore Brede, Christian Schjølberg, Martin Stige, Odd Erik Gundersen, and Morten Vinter in the Volve development team.

### References

- [1] Erik Olsen, Peter Funk, Ning Xiong. Fault diagnosis in industry using sensor readings and case-based reasoning. *Journal of Intelligent & Fuzzy Systems*, Vol. 15, ISSN 1064-1, p10, IOS Press, December, 2004.
- [2] Agnar Aamodt. Knowledge-intensive case-based reasoning in Creek. In Peter Funk, Pedro A. Gonzalez Calero (eds.), *Advances in case-based reasoning*. Lecture Notes in Artificial Intelligence, LNAI 3155, Springer, 2004. pgs. 1-15. ISBN0302-9743.
- [3] Aminul Islam, Pål Skalle. Improved efficiency and knowledge-based support of oil well drilling through case-based reasoning. In Proceedings of Intelligent Energy 2008. Society of Petroleum Engineers. SPE 111849.
- [4] Volve AS company web page: <http://www.volve.no>
- [5] Francisco J. Martín and Enric Plaza (2004), Ceaseless Case-Based Reasoning. In P.A. Gonzalez and P. Funk (Eds.) *Advances in Case-Based Reasoning*. Lecture Notes in Artificial Intelligence LNAI 3155, Springer, 2004. pgs. 287-301. ISBN0302-9743.