Proceedings of the Twenty-Ninth International
Florida Artificial Intelligence Research Society Conference

Evaluation of Explanations Extracted from Textual Reports

Gleb Sizov, Pinar Öztürk, Kerstin Bach
Norwegian University of Science and Technology (NTNU)
NO-7491 Trondheim, Norway

Abstract

Explanations play an important role in AI systems in general and case-based reasoning (CBR) in particular. They can be used for reasoning by the system itself or presented to the user to explain solutions proposed by the system. In our work we investigate the approach where causal explanations are automatically extracted from textual incident reports and reused in a CBR system for incident analysis. The focus of this paper is evaluation of such explanations. We propose an automatic evaluation measure based on the ability of explanations to provide an explicit connection between the problem description and the solution parts of a case.

Introduction

Explanation is a concepts that is intuitive to understand but hard to define formally. They are often used by humans to support decisions and actions. The most common type of explanations are causal, answering the why-question, e.g. "John will be late for dinner because he is stuck in traffic". Other common types of explanations are functional, e.g. "Bats have echolocation for navigation in the dark.", and intentional e.g. "Arthur bought new training shoes because he wants to start training." In general, to determine what is an appropriate explanation it is necessary to consider its goal as recognised by Leake et al. (Leake 1995) and Sormo et al. (Sørmo, Cassens, and Aamodt 2005).

In artificial intelligence (AI) systems, explanations have two roles: *internal* and *external* (Aamodt 1991). The internal ones are used in a reasoning process by the system itself while external explanations are targeted at users of the system to show how the system solves a problem. When the performance task of a system is to explain an anomalous situation, the constructed explanation can play the role of both an internal and an external explanation. Therefore a system with explanatory capabilities can win the trust of the user.

Schank proposed a case-based approach to explanation (Schank 1986). The basic idea of this approach is to store, index and retrieve "explanation patterns" (XPs), which are specific or generalized explanations of events. XPs can be tweaked to adapt them to new situations as elaborated in (Schank and Leake 1989). This approach was implemented in the SWALE system by (Kass, Leake, and Owens 1986), which can explain complex real-life problems such as the death of the racehorse Swale, who was successful in his career but died prematurely of the unknown reason. The main weakness of SWALE is that it relies on extensive knowledge engineering to construct domain-specific explanation patterns. Meta-AQUA (Cox and Ram 1999) also deals with story understanding, with a special focus on metacognition, relying on domain-specific top-down generated explanation patterns similar to SWALE.

The approach investigated in our work resolves the knowledge engineering burden by not relying on manually constructed explanation patterns but using free text documents as knowledge resources. The overall assumption of our approach is that text documents may contain explanations that can be extracted and used in a CBR system. Such documents are incident reports, judicial records or service reports. The main characteristic of the reports should be that they contain knowledge about how a problem was solved, so that we can provide computational support for reusing such documented knowledge.

In our previous work we described how to extract case-specific explanations from text and showed their usefulness in the case retrieval and adaptation (Sizov, Öztürk, and Styrak 2014; Sizov, Öztürk, and Aamodt 2015) processes. In this paper we focus on the role of explanation as a connector between the problem description and the solution parts of a case. Explanations allow to trace the reasoning from the problem description to the solution of a case. A good case explanation should not leave parts of the solution unexplained. This underlies the new evaluation criteria for causal explanations we propose in this paper.

The remainder of this paper is structured as follows. First, we introduce the application domain of safety investigation reports. Next, two sections describe how cases are generated from raw text of investigation reports. Then we present our evaluation, introducing a novel evaluation measure for case explanations. We wrap up with conclusions and future directions.

Application Domain

We use investigation reports from Transportation Safety Board of Canada. These reports document the analysis of transportation incidents including aviation, marine, railroad and pipeline incidents. They are the results of incident investigations conducted by domain experts and are reviewed for quality before publishing for public access. Each report is a 5-10 pages long containing summary, factual information, analysis and conclusion sections. Most of the explanations are contained in the analysis section, e.g.

> The oil that burned away did not return to the tank and, after a short time, the oil level became very low, causing the engine oil pump to cavitate and the engine oil pressure to fluctuate. Furthermore, since the oil did not return to the tank, the oil temperature did not change, or at least not significantly, and the pilot falsely deduced that the engine oil pressure gauge was displaying an incorrect indication.

Generally speaking, a causal analysis is essential for solving complex problems such as diagnosing a patient, investigating an accident or predicting the outcome of a legal case. Human analysts perform a root cause analysis (RCA) to answer why a certain problem occurred in the first place. A problem is characterised by an undesired outcome such as a failure, accident, defect or a dangerous situation etc. Causes are the events or conditions that lead to the undesired outcome. RCA goes beyond the proximate causes that immediately and directly precede the outcome and, aims to find the root causes that create the proximate causes which in turn ultimately leads to the outcome. The root causes are of vital importance for the prevention of the same problems to occur in the future. Since the incident reports document the analyst's reasoning from effects to causes, these explanations, if they can be automatically extracted from the reports, have the potential to be reused by CBR systems.

Mapping Reports to Cases

Investigation reports have similar organization with sections that can be defined either as summary, analysis or conclusion section. The summary section provides a brief description of the incident which we map to the problem description part of a case as shown in Figure 1. The analysis section explains why the incident happened identifying its root causes and contributing factors. Note that an explanation is not just a label, it is a coherent chain of states or events that links symptoms to the plausible causes. These causes are enumerated in the conclusion section of a report. The analysis and conclusion sections of the report are mapped to the explanation and conclusion parts of the case solution.

Extraction of Cases from Reports

The extraction process makes use of natural language processing components integrated into a pipeline shown in Figure 2. As shown in Figure 1, the problem description part of a case is represented with the Vector Space Model (VSM), which obtained by using standard information retrieval methods (Salton, Wong, and Yang 1975). The solution of a case is the explanatory graph represented with the Text Reasoning Graph (TRG) representation. TRG captures explanations of reasoning underlying the analysis in a graph

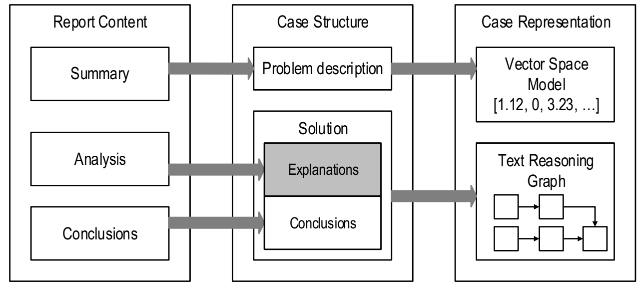

Figure 1: Mapping of textual reports to cases.

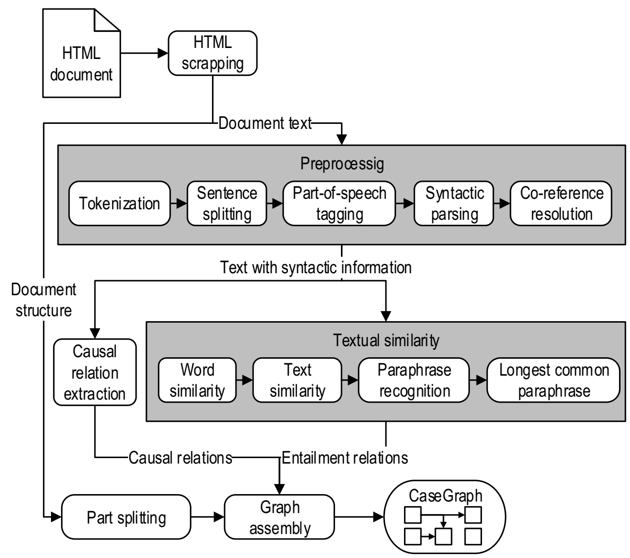

Figure 2: Pipeline for automatic acquisition of explanations from text: extraction of syntactic and semantic analysis of text in order to build CaseGraphs.

with nodes containing phrases and sentences and edges representing causal and entailment relations as shown in figure 3. It promotes understanding of why and how an incident has happened.

The process of extracting TRGs from text can roughly be divided into two phases: information extraction and graph assembly. In the information extraction phase, a report in the HTML format is converted into a structured text annotated with syntactic and semantic information. This information is then used in the graph assembly phase to generate a TRG. The following steps are included in the IE phase:

1. Parse the HTML of the report and extract text and sections.
2. Split the report into summary, analysis and conclusion parts based on the section titles, e.g. a section with the title containing the words “findings” or “causes” is assigned to the conclusion part. Similar lexical patterns were constructed for each part.
3. Process the report with the CoreNLP pipeline (Manning et al. 2014) (see Figure 2).

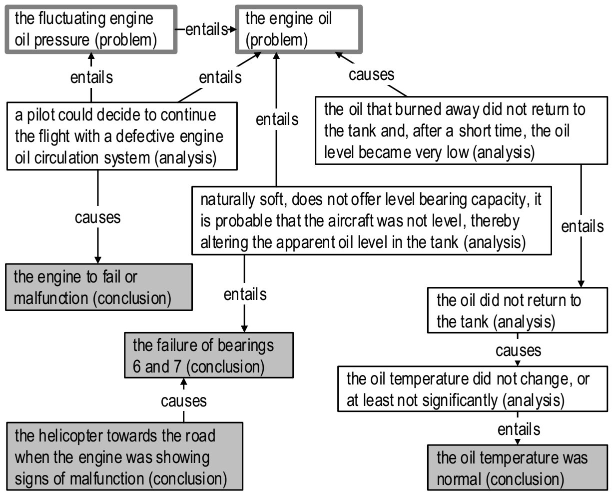

Figure 3: The text reasoning graph containing phrases and relations between them. The problem phrases are framed bold, the analysis phrases are in grey boxes and the conclusion in white boxes.

4. Extract causal relations from text of the report using patterns proposed by Khoo (Khoo 1995). These patterns are constructed around causal expressions such as "because of" or "as a result of" and incorporate phrasal categories e.g. NP - noun phrase, VP - verb phrase. For example the pattern "because of [CAUSE:NP], [EFFECT]" would match the sentence "Because of the durability of the coverings, it would be extremely difficult for a survivor with hand injuries to open the survival kit." and extract "the durability of the coverings" as the cause and "it would be extremely difficult for a survivor with hand injuries to open the survival kit" as the effect.

The graph assembly phase constructs a TRG from the extracted information following these steps:

1. Causal relations are collected in one graph with arguments as nodes and relations as edges.
2. Nodes that are not arguments of the same causal relation are connected by textual entailment relations to make the graph more connected. These relations are identified based on the textual similarity measure that uses WordNet to recognise words with similar meaning.
3. Nodes that are paraphrases of each other are merged into one node to remove redundant nodes with very similar meanings. The paraphrases are detected using the same textual similarity measure as for the textual entailment recognition.
4. Nodes with low informativeness are removed from the graph as described in the section below, e.g. the phrase "after a short time" does not carry much information by itself and is considered uninformative. Formally the informativeness is measured by computing the inverse document frequency (IDF) of the words in a phrase, which is commonly used in information retrieval for feature weighting.
5. As can be seen in Figure 3, the nodes are marked as problem, analysis, and conclusion nodes, according to which part in the report they are extracted from.

Dataset	Number of cases	Competence
Aviation	885	0.55
Marine	378	0.53
Rail	275	0.64

Table 1: Evaluation results, average competence of explanations

The details of each step are described in our previous work (Sizov, Öztürk, and Styrak 2014).

Evaluation

In our previous work we evaluated the use of textual explanations in retrieval and adaptation tasks, showing improvements compared to information retrieval baselines (Sizov, Öztürk, and Styrak 2014). In this work we evaluate the explanations as a connector between problem and solution parts within a case. A good explanation should be able to link anomalies in the problem description to conclusions in the solution. Based on this criteria we propose a novel evaluation measure called *explanation competence* that shows whether an explanation was able to explain the conclusions in terms of the problem description. Ideally all the conclusions should be explained.

For evaluation we use three report datasets from the Transportation Board of Canada collection including aviation, marine and rail incident reports. Each report contains a conclusion section that enumerates causes and contributing factors for the incident. We consider each cause and the factor in the conclusion section of a report as a separate conclusion and define our evaluation measure as the ratio between explained conclusions and the total number of conclusions in the corresponding investigation report:

$$competence(case) = \frac{|conclusion \in explanation|}{|conclusions|} \quad (1)$$

A conclusion is considered explained if it is connected to the problem description through the explanation chain in the TRG. Figure 4 shows examples of explanation chains that connect conclusions to pieces of information from the problem description. It is plausible to assume that all the conclusions in the report are explained. Otherwise the report would not be approved as lacking rigor. It means that a perfect explanation extraction system would achieve a competence score of 1. However, as shown in table 1, the competence scores obtained by our system are significantly lower, indicating that our explanation acquisition pipeline has room for improvement.

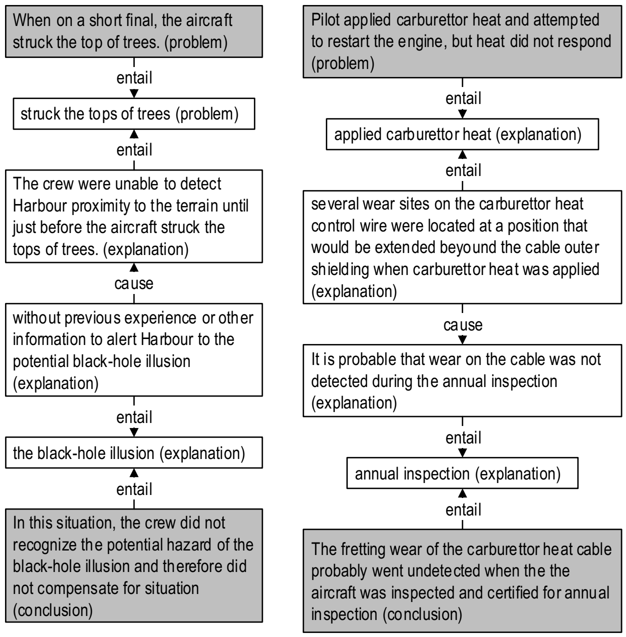

Figure 4: Example of explanations connecting problem (top nodes) to conclusions (bottom nodes)

The less than ideal scores can be attributed to errors in the NLP components. The causal relation extraction component is particularly important because causal relations are the core of the TRG representation. Missing relations lead to the broken links between the problem description and the conclusions. We conducted a separate evaluation of the causal relation extraction component on manually annotated 18 aviation reports with 410 relations. The obtained results, 64.39% precision, 51.85% recall and 56.61% f-score, indicate that the competence scores can be increased by improving the causal relation extraction component.

Conclusion and Outlook

We described how explanations contained in text can be extracted to a structured representation. For evaluation we proposed the explanation competence measure based on the ability of explanations to connect a problem description to conclusions in the solution. The results of the evaluation indicate that our explanation extraction pipeline has room for improvement, in particular the causal relation extraction component.

Automatically detecting and representing explanations in CBR will help users' understanding of complex problems as well as the suggested solution. The approach investigated in our research can be used in the domains where documents describing problem solving experiences are available. Our explanation extraction pipeline mostly relies on off-the-shelf NLP components that are available for many big natural languages such as English and German but might not be available for smaller languages.

References

Aamodt, A. 1991. A knowledge-intensive, integrated approach to problem solving and sustained learning. *Knowledge Engineering and Image Processing Group. University of Trondheim* 27–85.

Cox, M. T., and Ram, A. 1999. Introspective multistrategy learning: On the construction of learning strategies. *Artificial Intelligence* 112(12):1 – 55.

Kass, A.; Leake, D.; and Owens, C. 1986. Swale: A program that explains. *Explanation patterns: Understanding mechanically and creatively* 232–254.

Khoo, C. S.-G. 1995. *Automatic identification of causal relations in text and their use for improving precision in information retrieval*. Ph.D. Dissertation, The University of Arizona.

Leake, D. B. 1995. Abduction, experience, and goals: A model of everyday abductive explanation. *Journal of Experimental & Theoretical Artificial Intelligence* 7(4):407–428.

Manning, C. D.; Surdeanu, M.; Bauer, J.; Finkel, J.; Bethard, S. J.; and McClosky, D. 2014. The stanford corenlp natural language processing toolkit. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 55–60.

Salton, G.; Wong, A.; and Yang, C.-S. 1975. A vector space model for automatic indexing. *Communications of the ACM* 18(11):613–620.

Schank, R. C., and Leake, D. B. 1989. Creativity and learning in a case-based explainer. *Artificial Intelligence* 40(1):353–385.

Schank, R. C. 1986. Explanation: A first pass. *Experience, memory, and reasoning* 139–165.

Sizov, G.; Öztürk, P.; and Aamodt, A. 2015. Evidence-driven retrieval in textual cbr: Bridging the gap between retrieval and reuse. In Hüllermeier, E., and Minor, M., eds., *ICCBR-2015, Case-Based Reasoning Research and Development*, volume 9343 of *LNCS*, 351–365. Springer.

Sizov, G.; Öztürk, P.; and Styrak, J. 2014. Acquisition and reuse of reasoning knowledge from textual cases for automated analysis. In Lamontagne, L., and Plaza, E., eds., *ICCBR-2014, Case-Based Reasoning Research and Development*, volume 8765 of *LNCS*. Springer. 465–479.

Sørmo, F.; Cassens, J.; and Aamodt, A. 2005. Explanation in case-based reasoning–perspectives and goals. *Artificial Intelligence Review* 24(2):109–143.