Developing Mapping and Evaluation Techniques for Textual Case-Based Reasoning

Kevin D. Ashley and **Stefanie Brüninghaus**
Intelligent Systems Program and Learning Research and Development Center
University of Pittsburgh, Pittsburgh, PA 15260
ashley+@pitt.edu, steffi+@pitt.edu

Textual Case-Based Reasoning (CBR) is not simply Information Retrieval (IR) of text documents which happen also to be cases. Nor does it involve only techniques for automatically determining what cases represented as texts are about or techniques for automatically indexing such cases under relevant features.

Textual CBR is still case-based reasoning, and for us, that means drawing inferences about problem situations by comparing them to past decided cases. This specific property distinguishes textual CBR from other techniques such as IR. In other words, we believe that (1) textual CBR should also involve drawing inferences by computationally comparing cases expressed as texts. More accurately, given the technical difficulties of computationally comparing case texts, we take this as our working hypothesis. We believe it is worth investing a great deal of effort to confirm this hypothesis before we are willing to accept a watered-down conception of what textual CBR should be about.

At the same time, we do not minimize the difficulties of computationally comparing case texts. While cases may be compared computationally in terms of numerical feature weights, symbolic arguments, or adaptations, arranging for cases described as texts to be compared in these ways presents many research questions.

Chief among the problems to be solved is mapping text to comparison-enabling structures. We maintain that (2) to be successful, textual CBR will require textual descriptions of cases to be mapped onto structural representations which facilitate computationally comparing cases. It is important to realize that textual CBR does not avoid the need for structured case comparisons. While it may be true, as the Call for Participation asserts, that "many CBR applications now require the handling of only semi-structured or even full-text cases rather than the highly structured cases of more traditional CBR systems", textual CBR cannot dispense with structure in so far as it is necessary to support case comparison. In this paper, we elaborate one such structure called "factors" which facilitates comparing problems and cases. Our experiments in automatically classifying cases expressed as texts are aimed at identifying the factors implicit in the textual case descriptions.

Evaluation is a crucial step in any research project, and evaluating textual CBR systems is particularly challenging. We maintain that (3) an evaluation of textual CBR has to be designed carefully to disentangle the aspects to be assessed from other aspects of what is inevitably a complex system and to adopt appropriate standards for comparing system performance. Moreover, evaluation criteria for measuring success must augment those commonly applied to assess information retrieval of textual documents. The widely-used IR measures precision/recall do not capture all of the information required for assessing the performance of a textual CBR system. We develop this position on evaluation in a separate position paper.

Project Background

Our positions on these issues have been developed in the context of our preliminary work in an NSF-sponsored project entitled, "Adding Domain Knowledge to Inductive Learning Methods for Classifying Texts". Our project involves developing machine learning methods for automatically classifying and indexing texts describing new problems or cases. The goal is to develop a system to guide the process of assigning abstract factors to textual descriptions of legal cases. Factors are a kind of expert domain knowledge useful in legal argumentation and, as developed below, characteristic of other textual case-based reasoning domain tasks. The program takes as inputs the raw texts of legal opinions and assigns as outputs the applicable factors. The program's training instances are drawn from a corpus of legal opinions whose textual descriptions of cases have been represented manually in terms of factors (Aleven 1997; Aleven & Ashley 1997; Ashley & Aleven 1997). If we are successful, using domain knowledge to guide automatic text classification could integrate information retrieval, machine learning and AI knowledge-representation techniques, help scale up case-based reasoning systems, and alleviate the problem of assessing the relevance of texts in increasingly large on-line databases.

The project builds upon resources and know-how accumulated in previous research projects. We developed an expert model of case-based reasoning, which is the basis for CATO, an intelligent learning environment to teach law students argumentation with previous cases available as texts. The texts are legal opinions in which judges record their decisions and rationales for litigated disputes. We have compiled a corpus of full-text descriptions of 150 cases and a parallel abstract representation of some important aspects of those cases which capture their content and meaning. Our model of expert legal reasoning relates a set of factors, stereotypical factual strengths and weaknesses which tend to strengthen or weaken a legal claim, with the more abstract legal issues to which the factors are relevant. The evidence that factors apply to a given case are passages in the text of the opinions. CATO's Factor Hierarchy relates factors to more aggregated concepts and ultimately to legal issues raised by the legal claim. Together factors and the Factor Hierarchy enable CATO to generate examples of legal arguments and to provide some feedback on a students' work. We think that using the representation as guidance, a machine learning program trained on the corpus could learn to classify which factors and issues apply in new cases presented as raw texts.

In the initial phase of the work, we have: (1) Confirmed empirically that case similarity judgments based on CATO's factor model correspond to similarity of the associated texts as measured by a vector space model. (2) Formulated a set of hypotheses how including knowledge about the meaning of and relations among the factors and about their use in the case-based reasoning system can enhance learning and improve the performance of an induction classifier. (3) Examined in what ways background knowledge about the domain and a more-knowledge rich representation of text documents can be beneficial. (4) Generated initial classification results with purely inductive learning methods to establish a baseline for comparing the knowledge-based methods. (5) Begun a process of marking-up in the opinion texts passages (i.e., sentences) associated with the conclusion that a factor applies in the case (or not). We are analyzing those sentences to determine what kinds of classifiers would be able to identify such sentences in new opinions. We will test such classifiers empirically and then, if the results warrant, explore ways in which a program could automatically identify such classifiers from a training set of such sentences associated with each factor.

Beyond the initial phase, our basic methodology will be to start with some existing inductive and statistical learning algorithms and add different kinds of domain knowledge, assessing whether the algorithm's performance improves. We plan to add: (1) domain knowledge about factors and the legal issues to which they relate; (2) general information about the structure of legal opinions; (3) information about the statutes quoted in an opinion; (4) information about those cases cited in an opinion whose factors are known. We also explore (a) how to combine inductive and analytical techniques to deal with small numbers of training instances and (b) how best to combine successful inductive, statistical, and knowledge-based methods.

The mapping problem is hard for reasons that are likely to affect other domains of textual case-based reasoning. Factors are a crucial structure which enables the program, and facilitates human users, to compare and contrast cases, assess the relevance of their similarities and differences to the case at hand, and judge their implications for the current argument. Yet, the language of the opinions is complex, and the mere fact that an opinion discusses factors does not necessarily imply that those factors actually apply to the case. Such a problem can not be satisfactorily solved with currently available techniques for the following reasons:

- While current natural language processing techniques may be helpful, they are not adequate. Full text opinions of legal cases are a challengingly complex kind of text. As already noted, courts draw subtle distinctions in their use of language which are probably beyond the grasp of NLP approaches. Opinions may be long and their language complicated. In addition, as target concepts, factors are fairly abstract interpretive characterizations of a wide variety of factual phenomena appearing in cases.
- Merely looking for words or patterns related to a factor, as is done in information extraction research, will not suffice. Just because a factor is mentioned in an opinion does not mean that it actually applies in the case. An opinion may report that the plaintiff asserts certain specified facts or that he denies them. But the presence in such statements of the words describing the facts does not mean that the court has determined that the facts actually were or were not present; it simply discusses the respective issues. Similarly, the quotations of statutory texts alone do not imply that the court has determined that the corresponding factors actually apply in the case.
- Purely inductive learning methods (such as neural networks or decision trees) require too large a training set of examples. For a complex training task like finding factors in cases, the training data in our case base is certainly not sufficient when only induction is used. In addition, the fact that the target concepts are abstract interpretive characterizations means that they will not be ideal concepts to learn with a purely inductive approach.
- Although information retrieval techniques can retrieve opinions that contain words and concepts related to factors, they cannot discriminate among assertions that (1) use the same words but are not really about the factor, (2) assert that a factor is present in a case, (3) assert that a factor is not present in a case, (4) or say something else about the factor. IR "case representations" in terms of text feature vectors cannot make such distinctions. For instance, suppose one wants to retrieve cases that do not involve a nondisclosure agreement. IR techniques will screen out cases that mention "nondisclosure agreement" including a case where the court says plainly that the plaintiff did not obtain a nondisclosure agreement.

Mapping Texts of Cases Into Comparison Structures

As stated, we maintain that an important initial goal for textual case-based reasoning is to develop techniques for mapping textual descriptions of cases onto those structural representations which facilitate computationally comparing cases, and to test whether background knowledge which existing CBR models associate with such comparisons can facilitate the textual mapping. Information filtering, for instance, with its current preoccupation with automatically labeling subject matter of texts, will not suffice. Instead, textual CBR needs to address (1) how interesting or important a document is likely to be in light of a more detailed model of a reader's interests and of what use the reader can make of the document, (2) develop structural representations for comparing how cases relate to those models of users' interests, and (3) explore how textual case descriptions can be mapped onto those structural representations.

In our domain of legal argumentation, for instance, an opinion is more useful to the extent that it can provide support for a position the reader (a law student or attorney) wishes to maintain (or which he knows he/she must attack) concerning a problem scenario. If the opinion deals with a case in which the reader's position prevailed in similar or even weaker circumstances than those in the problem scenario, the opinion could be very valuable. Factors (i.e., stereotypical factual strengths and weaknesses) are useful structures for aligning the factual situations in the problem and case for purposes of comparing their strengths and weaknesses. As explained above, however, mapping opinion texts to factors in a case is a complex task even for a human reader, much less a computer program.

Given our database of case texts classified by applicable factors with specific sentences linked to individual factors, we are testing whether background knowledge can augment the performance of machine learning algorithms (such as Rocchio, TFIDF-Prototype, Winnow, Widrow-Huff, Libbow and Exponentiated-Gradient. See (Brüninghaus & Ashley 1997). We believe that background knowledge about factors can make up for the low numbers of classified examples. For instance, relations among factors and issues represented in the Factor Hierarchy may enhance performance by supporting predictions of the likelihood that a factor applies. The Factor Hierarchy expresses semantic relations among factors via more abstract parent concepts (i.e., legal issues and high-level factors). As the Factor Hierarchy represents: certain factors raise certain legal issues; some pairings of factors are inconsistent, or at least, highly unlikely; other pairings are quite common. These relations can provide confirming evidence for a factor in a text and can test the credibility of proposed factor assignments. Another source of classification information arises through the custom of citation. In writing legal opinions, frequently judges cite prior cases in support of their positions. Where the cases cited in an opinion are also in the database (as may often occur in a domain like trade secret misappropriation) , the classifier may use information about factors and issues present in the cited cases to help identify the factors and issues in the citing case.

In sum, if a mapping technique is to work realistically, we believe it will likely need to implement a model of the kinds of circumstances which are important to the reader, the conclusions to which they are relevant, and how the circumstances affect the conclusions in terms of relative strengths and weaknesses, benefits and costs, pluses and minuses. The most important information in a text may be the strengths and weaknesses, benefits and costs, or pluses and minuses of the subject matter it describes. This kind of information, which factors are intended to capture, may enable a program or a reader to compare the subject matter described to other cases known by or available to the reader. Factors are an example of one such model and structural representation. They relate low level factual circumstances described in the text to more specific conclusions about the text's usefulness in light of the reader's purpose.

While the details of the model, structural case comparison representations, document structures, and possibilities for mapping texts will differ across domains, other domains do evidence such models and structures. In a number of domains where case-based reasoning could be beneficial, one can construct a model of the use of documents and structures for comparing cases, the cases are available in the form of partially structured textual descriptions, and one can identify useful dimensions for comparing relative strengths and weaknesses, benefits and costs, pluses and minuses. For instance, large medical databases record patients' circumstances, diagnoses, prescribed treatments and outcomes in a somewhat structured form. Reuse and adaptation of this information is being attempted in such work as, for example, (Portinale & Torasso 1995). Perhaps abstract characterizations of the relative severity of medical conditions or invasiveness of medical procedures would be useful in comparing such case texts. Reuse and adaptation of software is also a potential CBR application, but one will need to match the software to the textual specification documents and to the handbooks and instructions created for describing the code. Comparisons in terms of complexity, cost, reliability or hardware requirements might be valuable in this domain. Detailed production plans and schedules can be adapted in a case-based process to new situations and reused, but realizing this in a practical way will require mapping general textual descriptions of production plans and constraints to the detailed structures. Here too, relative cost and complexity may be important factors for comparing textual cases. Help desk systems require free-text descriptions of customer service problems and solutions to be mapped and compared with other problems and solutions in terms, perhaps, of customers' degrees of sophistication or complexity of tasks.

It remains to be tested whether the texts in such task domains can be mapped to the comparison structures. As a reasonable starting point, we recommend exploring whether small sets of already classified cases, available background knowledge about comparing cases, and some information about document structure can serve to reduce the computational costs of mapping.

References

Aleven, V., and Ashley, K. 1997. Teaching Case-Based Argumentation through a Model and Examples: Empirical Evaluation of an Intelligent Learning Environment. In *Proceedings of the World Conference on Artificial Intelligence in Education (AI-ED 97)*, 87–94.

Aleven, V. 1997. *Teaching Case-Based Argumentation Through a Model and Examples*. Ph.D. Dissertation, University of Pittsburgh, Pittsburgh, PA.

Ashley, K., and Aleven, V. 1997. Reasoning Symbolically About Partially Matched Cases. In *Proceedings of the 15th International Joint Conference on Artificial Intelligence (IJCAI-97)*, 335–341.

Brüninghaus, S., and Ashley, K. 1997. Using Machine Learning for Assigning Indices to Textual Cases. In *Proceedings of the 2nd International Conference on Case-Based Reasoning (ICCBR-97)*, 303–314.

Portinale, L., and Torasso, P. 1995. ADaPTER: An Integrated Diagnostic System Combining Case-Based and Abductive Reasoning. In *Proceedings of the 1st International Conference on Case-Based Reasoning (ICCBR-95)*, 277–288.