Use Case: On Semi-Supervised Learning of Legal Semantics

L. Thorne McCarty
Department of Computer Science
Rutgers University
New Brunswick, New Jersey 08903
mccarty@cs.rutgers.edu

1. CURRENT RESEARCH

The goal of my current research is to construct domain-dependent legal knowledge representation languages either automatically or semi-automatically from unstructured legal texts.

The first step towards this goal was presented in my paper in ICAIL 2007:

> "Deep Semantic Interpretations of Legal Texts," *Proceedings of the Eleventh International Conference on Artificial Intelligence and Law*, pages 217–224 (ACM Press, 2007).[1]

There were two main contributions here. First, I showed that a "state-of-the-art statistical parser . . . can handle even the complex syntactic constructions of an appellate court judge" by presenting the output of Michael Collins' parser applied to the full text of the judicial opinions in 111 federal civil cases, comprising a total of 15,362 sentences. Second, and more important, in my opinion, I showed that "a semantic interpretation of the full text of a judicial opinion can be computed automatically from the output of the parser." The main technical contribution of my paper was the specification of a *quasi-logical form*, or *QLF*, to represent the semantic interpretation of a sentence, and a *definite clause grammar*, or *DCG*, to compute the correct quasi-logical form from the output of the parser. The *DCG* was constructed manually, but with automated assistance, and it stabilized at approximately 700 rules. The *QLF* was based loosely on my *Language for Legal Discourse (LLD)*, and it was intended to serve as an intermediate step towards the construction of the full representation of a legal case in *LLD*.

The second step was presented in my paper in ICAIL 2015:

> "How to Ground a Language for Legal Discourse in a Prototypical Perceptual Semantics," *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Law*, pages 89–98 (ACM Press, 2015).[2]

I have been told that this paper is hard to follow, so I have developed two new versions of the same material for different audiences. For lawyers, I have written a longer and gentler introduction (with the same title) in the *Michigan State Law Review*, 2016 MICH. ST. L. REV. 511 (2016).[3] For computer scientists and mathematicians, I have developed a slide show:

> "Probability, Geometry, Logic: A Triptych for a Learnable Knowledge Representation Language," presented at the CUNY Graduate Center in Manhattan on April 25, 2017.[4]

For AI and Law researchers, I would recommend reading the law review article and the slide show side by side.

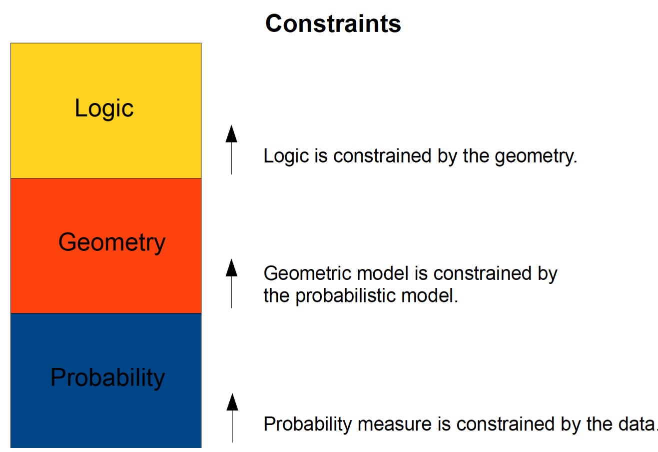

Figure 1: The Theory of Differential Similarity.

Figure 1 is a schematic view of the theory of *differential similarity*, as presented in these papers. The theory has several important properties:

1. It captures the "open-texture" of legal concepts, and it models (in its geometric layer) the "prototype-plus-deformation" structures that I have discussed in my previous work.

[1] Available online at `http://bit.ly/1Vk7gnk`.

[2] Available online at `http://bit.ly/1qCnLJq`.

[3] Available online at `http://bit.ly/2pkSfZl`.

[4] Available online at `http://bit.ly/2oVOAOH`. The mathematical prerequisites are: stochastic processes, differential geometry, category theory.

AILP Workshop at ICAIL 2017, June 12, 2017, London, England, UK

2. It combines a "rule-based" model and a "data-driven" model in a principled way, mediated by the geometric model in the center.

3. Although it is not a neural network, it implements a form of "deep learning," with multiple layers of unsupervised learning at the bottom and (optionally) a layer of supervised learning at the top.

4. Since the top layer is a logical language, the concepts that it learns are (potentially) interpretable.

Some aspects of this theory are still under development. For example, the language needs to be extended to encompass events/actions and the modalities over actions, such as permission and obligation, as well as the epistemic modalities of knowledge and belief, and several other common sense categories. But the papers cited above (and the references cited therein) outline how to do this.

Finally, we need to combine the work in my ICAIL 2007 paper with the theory in my ICAIL 2015 paper to learn a domain-dependent legal semantics from a corpus of unannotated legal texts, with only minimal supervision. This problem is discussed in my companion paper:

> "Discussion Paper: On Semi-Supervised Learning of Legal Semantics," *Workshop on Legal Text, Document and Corpus Analytics (LTDCA 2016)*, San Diego, California, USA (June, 2016).[5]

2. PROSPECTIVE APPLICATIONS

The basic "use case" for this research was presented almost 20 years ago in:

> "Structured Casenotes: How Publishers Can Add Value to Public Domain Legal Materials on the World Wide Web," *Second French-American Conference on Law and Computing*, Nice, France, 1998.[6]

The general argument in this paper was that editorial enhancements to primary legal materials (statutes, regulations, cases, etc.) should not take the form of additional natural language texts (treatises, annotations, practitioner's guides, etc.), but should take the form of *computational structures*, "using recent advances in Knowledge Representation (KR) and Natural Language (NL) techniques." Specifically, for the editorial enhancement of a legal case, I proposed the following definition:

> A structured casenote is a computational summary of the procedural history of a case along with the substantive legal conclusions articulated at each stage of the process. It would play the same role in the legal information systems of the 21st century that West Headnotes and Key Numbers have played in the 20th century.

[5]Available online at `http://bit.ly/2pBsZyO`.

[6]Available online at `http://bit.ly/1Trg2Qz`. A French translation of this paper by Danièle Bourcier was published as "L'Indexation de la Jurisprudence: Comment les Editeurs Peuvent Ajouter de la Valeur aux Données du Domaine Public sur le Web," in D. Bourcier, P. Hassett, and C. Roquilly, editors, *Droit et Intelligence Artificielle: Une Révolution de la Connaisance Juridique*, pages 191–200. Romillat, Paris, 2000.

The main body of the paper then explored this proposal by analyzing a recent copyright case, *Quality King Distributors, Inc., v. L'Anza Research International, Inc.,* 118 S.Ct. 1125 (1998), and its three opinions, in the District Court, the Court of Appeals for the Ninth Circuit, and the Supreme Court.

The focus on procedural history was based on the traditional "brief" that students are taught to write in their first year of law school. I explained the idea as follows:

> The traditional case brief focuses on the procedural context first: Who is suing whom, and for what? What is the plaintiff's legal theory? What facts does the plaintiff allege to support this theory? How does the defendant respond? How does the trial court dispose of the case? What is the basis of the appeal? What issues of law are presented to the appellate court? How does the appellate court resolve these issues, and with what justification?

To ask these questions and answer them in a structured casenote, I wrote, we need "a representation of the rules of civil procedure at some reasonable level of abstraction" (the KR component), and we need a computational grammar "with coverage of the procedural expressions that occur in the synopsis of a typical case" (the NL component). The structured casenote could then be elaborated as follows:

> Within this procedural framework, we would represent the substantive issues at stake in the decision. This is more complicated, since the territory is so vast, potentially encompassing the entire legal system. We can get started on this project, however, by focusing on particular areas of the law . . .

For example: We could focus on copyright law. The paper concluded by comparing the Supreme Court's decision in *Quality King v. L'Anza Research* with the prior decision in the Ninth Circuit, which turned on a disagreement about the scope of §§106(3), 109(a) and 602(a) of the 1976 Copyright Act. A structured casenote would represent this disagreement in a formal knowledge representation language (the KR component again), and link it back to the procedural history of the case. The paper showed how this would work.

There are obviously many other similar use cases, with different kinds of legal texts, different areas of the law, etc., but the pilot projects should be chosen carefully. The resources required for this project are substantial, both human resources and computational resources. My ICAIL 2007 paper relied on Michael Collins' parser, which was the best statistical parser available at the time. Today there are many alternatives. However, the *QLF* and the *DCG* are unique computational resources, which cannot be easily replicated. In my experiments with the theory of differential similarity, the computation of a geodesic coordinate curve takes about 90 minutes running on a desktop machine with a 3.06 GHz processor and 8 GB of RAM. But a simple problem might involve thousands of geodesic curves. The good news is that these computations are almost completely independent, and thus can be run efficiently on a massively parallel architecture, such as a GPU. This example gives some indication of the scale of the project.