Facts2Law – using deep learning to provide a legal qualification to a set of facts

Ivan Mokanov[†]
Lexum
Canada
ivan@lexum.com

Daniel Shane
Lexum
Canada
shaned@lexum.com

Benjamin Cerat
Lexum
Canada
ceratb@lexum.com

ABSTRACT

Over the course of the last year Lexum has started exploring the potential of deep learning (DL) and machine learning (ML) technologies for legal research. Although these projects are still under the umbrella of Lexum's research and development team (Lexum Lab, https://lexum.com/en/ailab/), concrete applications have recently started to become available. This demo focuses on one of these applications: Facts2Law.

The project benefits from a combination of factors. First, the millions of legal documents available in the CanLII database in parsable format along with structured metadata constitute a significant dataset to train AI algorithms. Second, Lexum has direct access to the knowledge and experience of one of the leading teams in AI and deep learning worldwide at the Montreal Institute for Learning Algorithms (MILA) of the University of Montreal. Third, the availability of computer engineers with cutting-edge expertise in the specifics of legal documents facilitates the transition from theory to practical applications.

Regarding concrete outcomes, Lexum's Facts2Law can predict the most relevant sources of law for any given piece of text (incorporating legal citations or not).

CanLII as a Training Data Set

CanLII is the largest repository of Canadian public legal information. The CanLII database includes:

- Data from 14 jurisdictions – federal, provinces and territories; More than 2M decisions from Canadian courts and tribunals. The length of these decisions varies greatly; their average is around ten pages;

- Around 650 statutes and a 2,000 regulations per jurisdiction, the vast majority of which have a table of contents represented as a tree whose leaves are numbered legislative "sections";

- A total of over 30,000 cited sections and subsections of statutes and regulations; Over 8.7M citations from one decision to another;

- Over 8M citations from decisions to a statute's section or subsection.

- All citations are hyperlinked and already extracted to a precision close to 90%;

- The citations and table of contents are encoded in a standardized fashion.

Lexum legal citator, Reflex, uses text analysis and probabilities to establish the associations between parallel citations. In the case of historical material, citation patterns from all major Canadian printed reports are supported. For more recent content, electronic citations patterns are supported (including the neutral citation, QL, Carswell, JE (Jurisprudence Express), Azimut, and REJB - EYB (Yvon Blais)). Reflex makes uses of parallel citations to expand the citation network of CanLII. For example, if Judge A cites the case X v. Y with its DLR1 and CCC2 citations, and Judge B cites the same case with its DLR1and OAC6 citations, Reflex will conclude that DLR1, CCC2, and OAC6 are all parallel citations of the same case. Reflex also use powerful computer heuristics – rules of thumb either manually coded or acquired through statistical analysis – to recognize oft-used citation patterns. Citations that are ambiguous by themselves can then be deciphered by clues given by their context. For example, if a full citation of the Immigration and Refugee Protection Act is detected at the beginning of a paragraph, a later mention of "the Act, at section ##" can be inferred to relate to the same statute.

Thanks to its citator, the CanLII database already includes a "map" of the Canadian Law. All of this data is already highly structured and available to train ML algorithms.

Facts2Law - Predicting Legal Citations

Lawyers regularly produce briefs, legal opinions, and other types of legal advice documents. These opinions, provided in writing, examine the various legal aspects of the client's situation. A lawyer who wants to research an aspect of that situation will often perform full-text search queries to buttress his opinion. Unfortunately, full-text search queries are often limited in their scope and might miss some nuances of the situation. To remedy

this, some systems provide search results based on the text of the whole opinion.

"More like this" systems are nothing new but they are typically bag of words affairs that are somewhat limited in their understanding of the content. In addition, legal databases are made up of words but also of citations, which are reliable indicators of popularity and authority, measures that are not considered in traditional "more like this" approaches.

Lexum's solution to this challenge is to learn from the existing citations on CanLII to predict which sources of law are relevant to the text of a legal brief, a legal opinion or to the plain language description of a legal issue.

This approach makes it possible to enhance content by providing additional contextual information. It also enables legal researchers to search the law in an entirely new way, by describing their problem in plain language. The results obtained will constitute a good starting point in the sorting of issues and the subsequent exploration of the applicable rules.

While there is no training data for this exact question in the context of the CanLII database, citations in case law can act as stand-ins for the question of relevance. First, court decisions are opinions with content and form very similar to legal opinions and briefs. Second, the citations that appear in these court decisions are typically the result of a research and relevance evaluation by a human being in the course of his research. If a document is cited, it is because it is relevant. If a document is not relevant enough to be discussed, then it will not be cited.

Lexum has developed a preliminary iteration of Facts2Law. It can currently be interrogated via a simple web-based interface. As a baseline control the current tool makes use of the Logistic Regression for ML (https://machinelearningmastery.com/logistic-regression-for-machine-learning/), a technique borrowed from the field of statistics.

The production instance makes use of a Neural Networks and Deep Learning algorithm.

This approach takes as input both a brief and a target document and, as output, its confidence that the brief should cite the target. To do so, Lexum uses whole document embeddings of both documents and a weighted summary of those in neighbouring nodes in the citation graph along with the relevant metadata. Then, the approach consists in feeding both branches to fully connected layers, merge them and modulate the merged inputs to take into account the age of the brief before passing them to another fully connected layer that outputs the prediction. This architecture allows us to use heuristics to select a subset of the corpus that could be relevant to a document and then rank them efficiently to extract documents of interest.

Example

The following capture from Facts2Law shows the prediction of Section 72 of the Immigration and Refugee Protection Act when queried with the following text found on the Web:

"I got a rejection letter today for my visa application (for a 6 week tourist visitor trip) to Canada from India. The reasons given, however, are very vague. None of the checkboxes for documents are ticked but it says under "Other reasons" (ticked):

"Given economic conditions, employment prospects, considering your travel history, economic establishment and family ties, I am not satisfied that you would respect the terms of your admission as a temporary resident in Canada".

This is quite unclear and surprising to me because I had applied to visit my cousin on holiday. I am still employed in India and submitted my passport (showing my travel history), flight bookings (for this trip), last 6 payslips, 4 months bank statements and tax returns, along with other required documents. Moreover, I had sent visa invitation letter, passport, work permit details of my cousin.

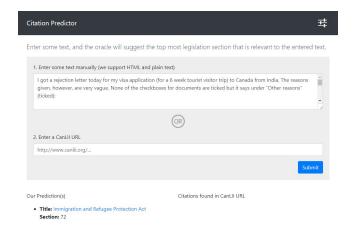Has anyone else received a similar rejection letter? Any suggestions on how to re-apply and address these rejection reasons would be very helpful."

Figure 1: Facts2Law - capture