



Natural Language Processing and Machine Learning Techniques in Real World (Law and Health)



Mi Young Kim*

Department of Science, Augustana Faculty, University of Alberta, Canada

Submission: September 08, 2017; Published: September 21, 2017

*Corresponding author: Department of Science, Augustana Faculty, University of Alberta, 4901-46 Avenue, Camrose, AB, T4V 2R3, Canada, Tel:+1780 679 1104; Email: miyoung2@ualberta.ca

Editorial

Every day a large volume of legal documents are produced, and lawyers need support for the analysis of the big documents, especially in corporate litigation. Typically corporate litigation has the aim of finding evidence for or against the litigation claims. Identifying the critical legal points from large volumes of legal text is time consuming and costly, but recent advances in Natural Language Processing and Machine Learning have provided new enthusiasm for improved management of legal texts and the identification of legal relationships. I believe that developing tools to automatically, or semi-automatically confirm entailment relationships between legal texts is fundamental to legal text understanding. Many researchers are interested in developing question-answering tools which help novice users obtain information about their questions in law, and an initial important step is the development of a question answering system for Yes/No questions.

Yes/No question answering is significantly easier than the general question answering, but we still need to develop tools for determining the relationships amongst legal texts. There has been a Competition on Legal Information Extraction/Entailment (COLIEE) Kim & Goebel [1] since 2014. The COLIEE Competition focuses on two aspects of legal information processing related to answering yes/no questions from legal bar exams: Legal document retrieval (Phase 1), and Yes/No Question answering for legal queries (Phase 2). Many systems have participated in the competition and showed promising results in Yes/No Question answering using Convolutional Neural Network Machine Learning techniques Kim et al. [2] and syntactic parsing of Natural Language Processing Marsi et al. [3], Haermeling [4].

I expect a general question answering system in law domain will be released soon with the improved Natural Language Processing and Machine Learning techniques. In the same way, medical domain has also needed computer science technologies

to extract information from a large volume of medical data. Most clinical data consist of unstructured natural sentences (e.g. clinical notes, Tele-Health transcripts), and there are few standard templates for the description format. Even existing templates are limited in scope, brittle in structure, and often adjusted so that calibration is difficult. It is also difficult to automatically extract that information which health care professionals believe is important to improve health care because there is much noise in most directly captured health data.

The development of a tool to parse patient records in order to automatically detect signs of a possible health issue would be a tremendous help for epidemiologists and other health professionals, and it could allow them to react more rapidly to a variety of trends. Recent advances in a variety of Artificial Intelligence (AI) Natural Language Processing techniques, such as information extraction, named entity recognition, and factual assessment, support the development of such tools. As an integral part of Electronic Health Records (EHR), clinical notes pose special challenges for analyzing EHRs due to their unstructured nature and substantial noise because they are written by health practitioners in real time, while talking with patients.

The noise can be divided into two types: First is explicit noise, such as spelling errors, abbreviations, unspecified acronyms, unfinished sentences, term variants, and omission of sentence delimiters. The second is implicit noise, which is revealed only by a variety of inference methods: We have to figure out the intended meaning and linguistic structure of sentences to detect implicit noise. Written information that is not about a patient and untrustworthy information that may not be true are examples of implicit noise.

As part of our noise identification process, we have to detect biomedical name entities-such as disease, virus, drug, and

symptom-as well as temporality, temperature, travel history, and other kinds of patient personal named entities (e.g., age and sex). Biomedical named entity recognition is the recognition of technical terms in the biomedical fields such as diseases, drugs, symptoms, and the like. Examples of biomedical named entity recognition systems include extracting clinical information from radiology reports Fiszman et al. [5], identifying diseases and drug names in discharge summaries Uzuner et al. [6], and detecting gene and protein mentions in biomedical paper abstracts Yeh et al. [7]

There are many systems which showed successful identification of biomedical named entities in clinical notes by removing the described noise using Natural Language Processing and Machine Learning techniques Kim et al. [8]. Furthermore, there are systems which automatically detect frailty of senior patients Castell et al. [9] and early predict pandemic diseases by analyzing Twitter texts Signorini et al. [10]. In near future, health professionals will rely on the Natural Language Processing and Machine Learning Techniques in analyzing their patient data, and system will even recommend proper treatment for each patient based following the analysis results.

Acknowledgement

This research was supported by the University of Alberta Start-up funds.

References

1. Kim MY, Goebel R (2017b) Two-step Cascaded Textual Entailment for Legal Bar Exam Question Answering. International Conference on Artificial Intelligence and Law. London, UK.
2. Kim MY, Xu Y, Goebel R (2017a) Applying a Convolutional Neural Network to Legal Question Answering. Lecture Notes in Artificial Intelligence 10091: 282-294.
3. Marsi E, Krahmer E, Bosma W (2007) Dependency-based paraphrasing for recognizing textual entailment. In Proceedings of ACL PASCAL Workshop on Textual Entailment and Paraphrasing 83-88, USA.
4. Harmeling S (2007) An extensible probabilistic transformation-based approach to the third recognizing textual entailment challenge. In Proceedings of ACL PASCAL Workshop on Textual Entailment and Paraphrasing, pp. 137-142.
5. Fiszman M, Chapman W, Aronsky D, Evans R, Haug P (2000) Automatic detection of acute bacterial pneumonia from chest X-ray reports. J Am Med Inform Assoc 7(6): 593-604.
6. Uzuner O, South B, Shen S, DuVall S (2010) i2b2/va challenge on concepts, assertions, and relations in clinical text. J Am Med Inform Assoc 18(5): 552-556.
7. Yeh A, Morgan A, Colosimo M, Hirschman L (2005) Bio creative task 1a: Gene mention finding evaluation. BMC Bioinformatics 6(Suppl 1): S2.
8. Kim MY, Xu Y, Zaiane OR, Goebel R (2015) Recognition of Patient-Related Named Entities in Noisy Tele-Health Texts. In ACM Transactions on Intelligent Systems and Technology TIST, Article No.59, New York, USA.
9. Castell MV, Sanchez M, Julian R, Queipo R, Martin S, et al. (2013) Frailty prevalence and slow walking speed in persons age 65 and older: implications for primary care. BMC Fam Pract 14: 86.
10. Signorini A, Segre AM, Polgreen PM (2011) The use of Twitter to track levels of disease activity and public concern in the US during the influenza A H1N1 pandemic. PLoS one 6(5): e19467.



This work is licensed under Creative Commons Attribution 4.0 License
DOI: [10.19080/RAEJ.2017.01.555562](https://doi.org/10.19080/RAEJ.2017.01.555562)

Your next submission with Juniper Publishers will reach you the below assets

- Quality Editorial service
- Swift Peer Review
- Reprints availability
- E-prints Service
- Manuscript Podcast for convenient understanding
- Global attainment for your research
- Manuscript accessibility in different formats
(Pdf, E-pub, Full Text, Audio)
- Unceasing customer service

Track the below URL for one-step submission
<https://juniperpublishers.com/online-submission.php>