Fast AI classification for analyzing construction accidents claims

Rita Yi Man Li
Department of Economics and Finance
Hong Kong Shue Yan University
Hong Kong
ymli@hksyu.edu

Herru Ching Yu Li
HKSYU Real Estate and Economics Research Lab
Hong Kong Shue Yan University
Hong Kong
herru24@gmail.com

Beiqi Tang
HKSYU Real Estate and Economics Research Lab
Hong Kong Shue Yan

WaiCheung Au
HKSYU Real Estate and Economics Research Lab
Hong Kong Shue Yan University
Hong Kong
auwaiceung01@icloud.com

ABSTRACT

Safety has long been considered an important issue in the construction industry. One means of reducing accidents is to provide for heavy compensation. As per the common law system, precedents, once made, then become part of the legal system. Therefore, construction companies and legal firms have an interest in obtaining details of court cases relevant to the ones they are currently involved with. However, the cost of identifying relevant court cases can be excessive. Computer-based text classification, the process of classifying documents into predefined categories with regard to their content, is proposed in this paper as a way to speed up the procedure whereby court cases are identified as relevant to a particular claim for accident compensation. The data set used for this project consisted of 3000 sentences. The 'training split' was 90% training and 10% for testing. The results show that the precision of the system proposed here for classifying construction accident cases is 95.7% and that the recall is 95.7%. This demonstrates that the fastText-based classification employed can successfully classify papers as relevant to the acceptance or rejection of a compensation case at a fairly high rate of accuracy. This pilot research provides a practical example in order to showcase the possibility of utilising artificial intelligence, without human intervention, for document classification. Such a facility could reduce the time taken to identify relevant past cases, so saving human resources, and improving turn-round times.

AICSconf '20, August 20–22, 2020, Wuhan, China
 ACM ISBN 978-1-4503-7727-0/20/08…$15.00
https://doi.org/10.1145/3407703.3407705

CCS CONCEPTS

• Information systems→Information retrieval→Specialized information retrieval→Structure and multilingual text search→Structured text search

KEYWORDS

Natural Language Processing, construction accident, classification, artificial intelligence

ACM Reference format:
Rita Yi Man Li, Herru Ching Yu Li, Beiqi Tang, Wai Cheung Au. 2020. Fast AI classification for analyzing construction accidents claims. In *Proceedings of 2020 Artificial Intelligence and Complex Systems Conference (AICSconf 2020). August, 2020, Wuhan, China, 4 pages.* https://doi.org/10.1145/3407703.3407705

1 INTRODUCTION

The construction industry is a major player in terms of the economic development of many countries. Compared to other industries, the construction industry has recorded the highest rate of non-fatal injuries or illnesses requiring days off work [1]. Ensuring the health and safety of workers on site is of utmost importance in relation to construction and building processes [2]. In recent years, most scientific work devoted to the problem of engineering and construction processes has included research on the health and safety of construction workers. Li and Poon [3] suggested that construction workers in Hong Kong have often faced serious injury because of the lack of knowledge sharing. Szóstak [4] recognised that workers and participants are exposed to a number of risk factors relating to injury and death. Li [5] investigated construction safety informatics including on-site AI, chatbots, and the IoT, and their possible roles in reducing the occurrence of construction accidents.

When filing a building accident lawsuit, Murphy [6] suggests that it is important to work with an experienced attorney who understands what ultimately affects how much you may receive for your injuries. Setting the right compensation level for injured workers can be very complex. Multiple parties, including the employer and other parties involved in the construction project, may all be held liable for an accident. Barsumian [7] believed that Injury attorneys could file lawsuits against negligent makers/owners/maintainers of equipment, property owners who did not disclose their greater knowledge of hazards, and other construction companies who had not adhered to best safety practices on construction sites. Hiring an experienced personal injury lawyer who knows how to identify the causes of an injury, such as negligence, is the first step in getting the compensation you deserve. Gürcanli and Müngen [8] indicated that if a person's negligence leads to an avoidable building accident, that person should be held liable for the injury to the victim. It is always important to look at who is to blame for building accidents. A safety expert is responsible for determining whether there is a link between the safety standards of a construction site and the risk of injury of accident victims. Sometimes, the architect and/or engineer may be liable. Miller [9] determined that in some construction accident court cases, the system architect and/or the engineer responsible for designing systems supporting the design, construction, maintenance and operation of buildings could be held liable for their negligence or improper use of materials/equipment.

2 LITERATURE REVIEW OF DOCUMENT CLASSIFICATION BY NATURAL LANGUAGE PROCESSING

Document classification by natural language processing has been discussed from a number of different perspectives. "NLP" is a generic term that includes techniques that take into account the complexity of language [10]. NLP represents a huge leap forward in machine learning and natural language processing [11]. The range of applications related to the field seems to increase every day, and NLP can be used to facilitate the undertaking of many tasks. Sinoara, et al. [12] believed that for the achievement of NLP, there was a number of methods and algorithms that could be applied, such as machine leaning. Barba [13] suggested that machine learning as applied to NLP (text analytics) involved, centrally, identifying the most relevant words and phrases in a text, such as "hello," "good" and "bad." This technique can be expressed as a model as applied to a data set. Such approaches are examples of supervised machine learning. NLP may also be implemented as an algorithm that works with large data sets to extract meaning; such methodologies are known as as unattended machine learning. Steven Bird [14] indicated that there was often a need to solve, effectively simultaneously, a number of classification problems which were closely linked. To obtain the dependencies between related classification tasks, we can apply common classifiers and models that choose a suitable label for each collection and related input. Sequence classification is another methodology. One sequence classification strategy is known as sequential classification or greedy sequence classification. This proceeds by trying to find the best labels for the next input and then applying the answer to find a good label for the next (plus 1) input; labels may be anything, "good" and "bad" are often used. In terms of applying NLP to document classification, Biswas [15] argued that there were a lot of NLP tools that could help, such as MALLET and Apache Open NLP.

3 LITERATURE REVIEW OF ARTIFICIAL INTELLIGENCE FOR DOCUMENT CLASSIFICATION

Recently, more and more people have worked to shed light on Artificial Intelligence. Li [16] indicated that the application of AI can be critical to the reduction of construction accident risks. Li and Leung [17] undertook research into finding ways of enabling AI to detect hazards on construction sites. Many experts have pointed out that AI can be helpful for document classification. M. Thangaraj [18] argued that AI could play an important role in the text mining arena. Kamruzzaman [19] stated that text classification was the process of classifying documents into predefined categories based on their content. The process should assign the natural language text to the various different categories automatically. Text classification is a primary requirement of text retrieval systems --- which retrieve texts in response to a user query. Such a system may transform the text in some way such as producing summaries, answering questions or extracting data. Linss [20] describes an AI based tool which was able to classify results according to their relevance, determine how often an essay (n this case) had been cited, and display a visual representation of the relationships between studies. Mahoney, et al. [21] determined that the application of AI text classification could make it much more efficient as it could cull massive volumes of data and identify responsive documents for use in these matters.

4 CLASSIFYING LEGAL DOCUMENT USING ARTIFICIAL INTELLIGENCE

In this research, we have utilised fastText, a system developed by Facebook, for the classification of legal documents relating to compensation payments. The fastText system has a trained English language vector model derived from Wikipedia, Common Crawl and other sources. Every word is stored along with a number of different tokens. The morphology and word representations are replaced with vectors. Unlike word2vec and other word vector models, FastText has implemented n-grams of characters [22]. Words can be used easily to discover unknown vector representations due to the fact that the length, n, is variable. The model can also be adapted to both supervised and unsupervised learning. In this paper, we have implemented supervised learning: each sentence was represented as a bag of-grams and was normalized to keep the local word order the same [23]. In this paper, the author uses 3000 sentences as data. The training split is 90% for training and 10% for testing. The results show that the precision of the system proposed here for classifying construction accident

cases is 95.7% and that the recall is 95.7%. This demonstrates that a fastText classification can successfully classify papers as relevant to the acceptance or rejection of a compensation case at a fairly high rate of accuracy.

Table 1. Statistics table

Sentence in document	Confidence Percentage	Document Name
''G. QUANTUM OF DAMAGES The plaintiff has received employees' compensation in the sum of $833,322.51.''	98.06%	*Cheung Chung v. Softtrans Supply Chain [2018]*
''PSLA was awarded at $50,000 for the first accident and $150,000 for the second one.''	85.99%	*Cheung Chung v. Softtrans Supply Chain [2018]*
''PSLA was agreed at $150,000 between the parties''	90.90%	*Cheung Chung v. Softtrans Supply Chain [2018]*
''The Court awarded $84,000 as PSLA.''	99.97%	*Cheung Chung v. Softtrans Supply Chain [2018]*
''I would award $100,000, which is in my view reasonable, under this head.''	99.42%	*Cheung Chung v. Softtrans Supply Chain [2018]*
''The defendant accepts $5,000 to be awarded under this head.''	99.93%	*Cheung Chung v. Softtrans Supply Chain [2018]*
''What I would do here is simply to retain the sum of $3,000 to cater for all the extracurricular activities of T, and this sum would include Olympiad Math as well.''	87.34%	ESBU v LCH [2013]
''However, as T will be travelling during peak season, I think the sum of $19,000 including the air fares and all other expenses is about right.''	95.75%	ESBU v LCH [2013]
''As I said in paragraph 39 above, I would allow a monthly sum of $3,000 for all of T extracurricular activities, including the Olympiad Math which will be paid by the mother in the future.''	97.31%	ESBU v LCH [2013]
''I accept this being the actually monthly expenses of T in this item. Medical Expenses at $2,100[31] The wife now claims that T monthly medical expenses amount to $2,100 which is made up by the sum of $200 as general medical expenses and $1,900 being consultation fee with a psychologist called Annie Ho.[32] For general medical and dental expenses, I understand that both the mother and the father employers would provide medical coverage for their employees and children.''	74.40%	ESBU v LCH [2013]
''As to quantum, I propose to assess the further costs summarily at $2,500.''	98.13%	ESBU v LCH [2013]
''Out of this sum, it would be reasonable to attribute a sum of $2,000 being the father personal travelling expenses, and the balance of $3,000 being travelling expenses under the general expenses for the whole family.''	51.64%	ESBU v LCH [2013]
''For the period April 2000 to March 2003 I assess the expenses as $841,475.00 per annum and net income thus $486,289.00 per annum before tax, for which I shall allow a deduction of 16%.''	70.81%	Young vs Lee Chu [2003]
''The plaintiff has claimed, and the defendant conceded, an amount of $281,750.44 under this head, which I award.''	99.52%	Young vs Lee Chu [2003]
''The defendant has offered an amount of $10,000.00: if he is prepared to pay that sum, I am prepared to order.''	84.98%	Young vs Lee Chu [2003]
''Once again the defendant has offered the sum of $10,000.00 which appears reasonable to me.''	98.52%	Young vs Lee Chu [2003]

5 CONCLUSION

This research aims to explore the possibility of using fasText AI, a system developed by Facebook, to classify construction accident court cases involving compensation. The pilot test demonstrates that AI classification is able to provide highly accurate results. This would benefit any law firm whose work involves scanning a great many documents in order to identify relevant court cases, and also

construction companies when they have a need to search for compensation cases with specific attributes, and academic researchers undertaking compensation research. Further research is needed and might provide very useful results.

ACKNOWLEDGEMENT

The work described in this paper was supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China, Project no. UGC/FDS15/E01/18

REFERENCES

[1] J. H. John A. Gambatese, Michael Behm, "Investigation of the Viability of Designing for Safety," University of Florida , East Carolina University , Oregon State University, 2005, vol. 29. [Online]. Available: http://elcosh.org/document/1892/d000693/investigation-of-the-viability-of-designing-for-safety.html

[2] R. Y. M. Li, D. C. W. Ho, and B. Tang, "Factors Which Affect Construction Safety in Different Types of Construction Work," in *Advances in Safety Management and Human Factors*, Cham, P. Arezes, Ed., 2018// 2018: Springer International Publishing, pp. 67-76.

[3] R. Y. M. Li and S. W. Poon, "Future motivation in construction safety knowledge sharing by means of information technology in Hong Kong," *Journal of Applied Economic Sciences,* vol. 4, no. 9, 2009.

[4] M. Szóstak, "Analysis of occupational accidents in the construction industry with regards to selected time parameters," (in English), vol. 9, no. 1, p. 312, 2019, doi: https://doi.org/10.1515/eng-2019-0027.

[5] R. Y. M. Li, *Construction Safety Informatics*. 2019.

[6] B. O. Murphy. "NYC Construction Accident Lawyers on Your Side." https://www.blockotoole.com/Construction-Accident-Help/ (accessed 28 April, 2020).

[7] Barsumian. "Personal Injury Lawyers." https://www.barsumianlaw.com (accessed 28 April, 2020).

[8] G. E. Gürcanli and U. Müngen, "Analysis of construction accidents in Turkey and responsible parties," (in eng), *Industrial health,* vol. 51, no. 6, pp. 581-595, 2013, doi: 10.2486/indhealth.2012-0139.

[9] Miller. "Accident Lawyer." https://atlantaadvocate.com/georgia/atlanta/construction-accident-lawyer/ (accessed 28 April, 2020).

[10] P.-H. Chen, "Essential Elements of Natural Language Processing: What the Radiologist Should Know," *Academic Radiology,* vol. 27, no. 1, pp. 6-12, 2020, doi: 10.1016/j.acra.2019.08.010.

[11] D. L. Yse. "Your Guide to Natural Language Processing (NLP)." https://www.kdnuggets.com/2019/05/guide-natural-language-processing-nlp.html (accessed 28 April, 2020).

[12] R. A. Sinoara, J. Antunes, and S. O. Rezende, "Text mining and semantics: a systematic mapping study," *Journal of the Brazilian Computer Society,* vol. 23, no. 1, p. 9, 2017/06/29 2017, doi: 10.1186/s13173-017-0058-7.

[13] P. Barba. "Machine Learning for Natural Language Processing." https://www.lexalytics.com/lexablog/machine-learning-natural-language-processing (accessed.

[14] E. K. a. E. L. Steven Bird, "Learning to Classify Text," in *Natural Language Processing with Python*, 2019, ch. 6.

[15] S. Biswas. "Let's Talk About Natural Language Processing." https://dzone.com/articles/lets-talk-about-natural-language-processing (accessed 28 April, 2020).

[16] R. Y. M. Li, *An economic analysis on automated construction safety*. Springer, 2018.

[17] R. Y. M. Li and T. Leung, "Computer vision and hybrid reality for construction safety risks: a pilot study," *Advances in Intelligent Systems and Computing,* vol. 1027, 09/18 2019.

[18] M. S. M. Thangaraj, "Text Classification Techniques: A Literature Review," *Interdisciplinary Journal of Information, Knowledge, and Management,* vol. 13, pp. 117-135, 2018, doi: https://doi.org/10.28945/4066.

[19] S. M. Kamruzzaman, "Text Classification using Artificial Intelligence," *Journal of Electrical Engineering,* vol. 33, pp. 1-6, 2020.

[20] M.-A. Linss. "Looking Ahead: The Potential for AI in Literature Search." https://info.reprintsdesk.com/about/blog/2019/looking-ahead-the-potential-for-ai-in-literature-search (accessed 28 April, 2020).

[21] C. J. Mahoney, J. Zhang, N. Huber-Fliflet, P. Gronvall, and H. Zhao, "A Framework for Explainable Text Classification in Legal Document Review," presented at the 2019 IEEE International Conference on Big Data (Big Data), 2019.

[22] A. Agibetov, K. Blagec, H. Xu, and M. Samwald, "Fast and scalable neural embedding models for biomedical sentence classification," *BMC Bioinformatics,* vol. 19, no. 1, p. 541, 2018/12/22 2018, doi: 10.1186/s12859-018-2496-4.

[23] B. Kuyumcu, C. Aksakalli, and S. Delil, "An automated new approach in fast text classification (fastText): A case study for Turkish text classification without pre-processing," in *3rd International Conference on Natural Language Processing and Information Retrieval*, Tokushima, 2019.