

Text mining (also referred to as text analytics) is an artificial intelligence (AI) technology that uses natural language processing (NLP) to transform the free (unstructured) text in documents and databases into normalized, structured data suitable for analysis or to drive machine learning (ML) algorithms.

Text mining with automatic extraction of key features is gaining increasing importance in science and particularly medicine due to the rapidly increasing number of **publications**.

Objectives: Here we evaluate the current potential of **sentiment analysis** and machine learning to extract the importance of the reported results and conclusions of randomized trials on stroke.

Methods: PubMed abstracts of 200 recent reports of randomized trials were reviewed and manually classified according to the estimated importance of the studies. Importance of the papers was classified as "game changer", "suggestive", "maybe" "negative result". Algorithmic sentiment analysis was subsequently used on both the "Results" and the "Conclusions" paragraphs, resulting in a numerical output for polarity and subjectivity. The result of the human assessment was then compared to polarity and subjectivity. In addition, a neural network using the Keras platform built on Tensorflow and Python was trained to map the "Results" and "Conclusions" to the dichotomized human assessment (1: "game changer" or "suggestive"; 0: "maybe" or "negative", or no results reported). 120 abstracts were used as the training set and 80 as the test set.

Results: 9 out of the 200 reports were classified manually as "game changer", 40 as "suggestive", 73 as "maybe" and 32 as "negative"; 46 abstracts did not contain any results. Polarity was generally higher for the "Conclusions" than for the "Results". Polarity was highest for the "Conclusions" classified as "suggestive". Subjectivity was also higher in the classes "suggestive" and "maybe" than in the classes "game changer" and "negative". The trained neural network provided a correct dichotomized output with an accuracy of 71% based on the "Results" and 73% based on "Conclusions".

Conclusions: Current statistical approaches to text analysis can grasp the impact of scientific medical abstracts to a certain degree. Sentiment analysis showed that mediocre results are apparently written in more enthusiastic words than clearly positive or negative results ¹⁾.

1)

Fischer I, Steiger HJ. Toward automatic evaluation of medical abstracts: The current value of sentiment analysis and machine learning for classification of the importance of PubMed abstracts of randomized trials for stroke. J Stroke Cerebrovasc Dis. 2020 Sep;29(9):105042. doi: 10.1016/j.jstrokecerebrovasdis.2020.105042. Epub 2020 Jun 23. PMID: 32807454.

From:
<https://neurosurgerywiki.com/wiki/> - **Neurosurgery Wiki**

Permanent link:
https://neurosurgerywiki.com/wiki/doku.php?id=text_mining

Last update: **2024/06/07 02:54**

