Text Mining Past Medical History in Electrophysiological Studies

Authors : Roni Ramon-Gonen, Amir Dori, Shahar Shelly

Abstract : Background and objectives: Healthcare professionals produce abundant textual information in their daily clinical practice. The extraction of insights from all the gathered information, mainly unstructured and lacking in normalization, is one of the major challenges in computational medicine. In this respect, text mining assembles different techniques to derive valuable insights from unstructured textual data, so it has led to being especially relevant in Medicine. Neurological patient's history allows the clinician to define the patient's symptoms and along with the result of the nerve conduction study (NCS) and electromyography (EMG) test, assists in formulating a differential diagnosis. Past medical history (PMH) helps to direct the latter. In this study, we aimed to identify relevant PMH, understand which PMHs are common among patients in the referral cohort and documented by the medical staff, and examine the differences by sex and age in a large cohort based on textual format notes. Methods: We retrospectively identified all patients with abnormal NCS between May 2016 to February 2022. Age, gender, and all NCS attributes reports were recorded, including the summary text. All patients' histories were extracted from the text report by a query. Basic text cleansing and data preparation were performed, as well as lemmatization. Very popular words (like 'left' and 'right') were deleted. Several words were replaced with their abbreviations. A bag of words approach was used to perform the analyses. Different visualizations which are common in text analysis, were created to easily grasp the results. Results: We identified 5282 unique patients. Three thousand and five (57%) patients had documented PMH. Of which 60.4% (n=1817) were males. The total median age was 62 years (range 0.12 – 97.2 years), and the majority of patients (83%) presented after the age of forty years. The top two documented medical histories were diabetes mellitus (DM) and surgery. DM was observed in 16.3% of the patients, and surgery at 15.4%. Other frequent patient histories (among the top 20) were fracture, cancer (ca), motor vehicle accident (MVA), leg, lumbar, discopathy, back and carpal tunnel release (CTR). When separating the data by sex, we can see that DM and MVA are more frequent among males, while cancer and CTR are less frequent. On the other hand, the top medical history in females was surgery and, after that, DM. Other frequent histories among females are breast cancer, fractures, and CTR. In the younger population (ages 18 to 26), the frequent PMH were surgery, fractures, trauma, and MVA. Discussion: By applying text mining approaches to unstructured data, we were able to better understand which medical histories are more relevant in these circumstances and, in addition, gain additional insights regarding sex and age differences. These insights might help to collect epidemiological demographical data as well as raise new hypotheses. One limitation of this work is that each clinician might use different words or abbreviations to describe the same condition, and therefore using a coding system can be beneficial.

Keywords : abnormal studies, healthcare analytics, medical history, nerve conduction studies, text mining, textual analysis