Natural Language Processing Analysis based on Clinical Free Text Data using Computing Algorithm

Jihwan Park*, Mi Jung Rho**, Hyong Woo Moon*, and Ji Youl Lee*
*Department of Urology, Seoul St. Mary's Hospital, College of Medicine, the Catholic University of Korea, Seoul, 06591 Republic of Korea
**Catholic Cancer Research Institute, College of Medicine, The Catholic University of Korea, Seoul, 06591 Republic of Korea
e-mail:bosoagalaxy@gmail.com

의료 진단문 데이터 기반의 컴퓨팅 알고리즘을 이용한 자연어 처리 분석

박지환*, 노미정**, 문형우*, 이지열*
*가톨릭대학교 서울성모병원 비뇨기과 연구실
**가톨릭대학교 암연구소

요 약

Backgrounds: Clinical free text, from electronic medical records, consists of many medical terms which are not standard format. However, it has many important information that can help to assist clinicians decision making. Therefore we try to extract and analyze information from the clinical free text.
Results: We found there are many different terms which have same meaning. These terms can be analyzed by natural language processing with deep learning method. The analysis computing algorithm can extract terms by frequencies. This algorithm also found patterns in the clinical free text.
Conclusion: The analysis computing algorithm extract information from cumbersome free text from electronic medical records. This result can be expended many clinical site to find meaningful information without manual works.

Keyword: Clinical Free Text, Electronic Health Records, Natural Language Processing

1. Introduction

We can easily found clinical free text from electronic health records. Clinical free text consists of many medical terms and numbers from laboratory result[1]. Though it has many information, its contents are not standard. For instance, it has many different terms which have same meaning. When this issue were solved, we can find many important information that can assist clinicians decision making. That is why we try to extract information from clinical free text with natural language processing.

2. Materials and Methods

From university hospitals, we obtained clinical free data of prostate cancer. This study was approved by the institutional review board of the participating institutes. The data comprised several temporal clinical records including laboratory results, surgery, medication, and treatment. We used R language to develop computing algorithm for extracting information from clinical free text. R language is common method of development of clinical decision support system[2].

3. Results

We developed computing algorithm to extract information from clinical free text.
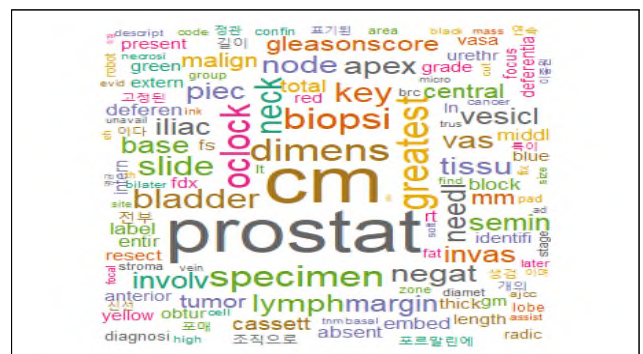
[Fig. 1] Analysis Example from Clinical Free Text

We successfully extract terms from clinical free text (Fig. 1). Based on this result we can manage the format of the free text. For instance, we can get needle biopsy result by searching terms such as needle and biopsy. Getting these information from clinical free text followed by analysis of an target medication or surgery.

4. Discussion

Though, clinical free text has many information it can not be the source of input of analysis, unless it has transferred to meaningful information. The initial result of this study shows an possibilities of finding useful information from cumbersome text data which as complex terms and numbers. Using the natural language processing result, of clinical free text, can be used to analyze certain disease[3].

5. Conclusion

We developed natural language processing computing algorithm for clinical free text. This result can be used to extract information from massive clinical data which exists in many hospitals.

참고문헌

[1] Sheikhalishahi S, Miotto R, Dudley JT, Lavelli A, Rinaldi F, Osmani V. Natural Language Processing of Clinical Notes on Chronic Diseases: Systematic Review. JMIR Med Inform 2019;7(2):e12239

[2] Park, J.; Rho, M.J.; Dritschilo, A.; Choi, I.Y.; Mun, S.K. Prostate Clinical Outlook Visualization System for Patients and Clinicians Considering Cyberknife Treatment-A Personalized Approach. Appl. Sci. 2018, 8, 471.

[3] W.W. Chapman, L.M. Christensen, M.M. Wagner, P.J. Haug, O. Ivanov, J.N. Dowling, et al. Classifying free-text triage chief complaints into syndromic categories with natural language processing Artif. Intell. Med., 33 (1) (2005), pp. 31-40