

4th International Conference and Exhibition on

Pharmacognosy, Phytochemistry & Natural Products

August 29-31, 2016 Sao Paulo, Brazil

Clinical effect extraction from Eastern medicine database

Jaehyun Lee¹, Yeeok Kang^{1, 2} and Doheon Lee¹¹Korea Advanced Institute of Science and Technology, Republic of Korea²SD Genomics Co., Ltd., Republic of Korea

Traditional Chinese medicine (TCM) has empirically accumulated knowledge for over 2,000 years. Traditional Chinese medicines integrated database (TCMID) is a large-scale TCM database with the unstructured plain text describing the functions of herbs and formulae. To analyze the high dimensional data using computational approaches, the plain text should be structured by recognizing the keywords (clinical entities and effect entities) and pair clinical-effect entities. This paper presents a pilot study and its result to extract specific and atomic clinical effects from the plain text based on the machine-learning approach. The main task was divided into two independent steps which were defined as supervised learning problems: effect entity detection and clinical-effect entity pairing. 100 herb entries and their functional descriptions from TCMID were randomly selected to generate the training corpus by manually tagging clinical effects. For the preprocessing step, MetaMap and BLLIP parser were utilized. MetaMap which is a tool for recognizing UMLS concepts was to recognize clinical entities with the semantic type filter. And then, BLLIP parser identified the deep-parsed structures and extracted syntactic features of the corpus for support vector machine (SVM) modeling. Based on the feature set, the two SVM classifiers were learned to detect the effect entities and to pair the clinical-effect entities, respectively. The proposed pipeline achieved an F-score of 88.97% on the eventual task. Therefore, clinical effect extraction to organize plain text in TCMID would promise the time- and cost-saving approach for drug developers to analyze TMC databases in an automated manner.

Biography

Jaehyun Lee has received his MS degree in Bio and Brain Engineering from Korea Advanced Institute of Science and Technology (KAIST), Korea in 2014 and is currently pursuing PhD. His research interests include bioinformatics, text-mining and machine learning.

jaeh@kaist.ac.kr

Notes: