Research on Potential Drug Side Effects Recognition Based on SAO Semantic Network

INTRODUCTION

With the continuous development of medicine and biotechnology, the therapeutic drugs for various diseases are becoming more abundant. According to the process of new drug development in the United States, new drugs need to undergo three stages of clinical testing before they go on the market. One of the purposes of clinical testing is to identify the efficacy and side effects to ensure the stability of the drug and to determine the side effects of the drug[1]. Therefore, it is particularly important to identify the potential side effects of drugs in advance in order to better adjust the subjects and methods of administration of new drugs, and to identify and predict the side effects of new drugs.

However, with the emergence of new drugs, the unknown adverse drug reactions will also increase, and a large number of random drug side-effect tests will bring unnecessary research and development costs. Therefore, predicting and identifying possible drug side effects plays a guiding role, which can save time and capital costs for research and development.

At present, the prediction of drug side effects based on text mining is a relatively common method. The text mining based on keyword has received a lot of attention from scholars, but this method can't effectively reflect the semantic relationship between medical entities. Moreover, the Subject-Action-Object (SAO) semantic analysis method can deeply explore the relationship between medical entities and can make up for the defects of keyword text mining. However, few articles use it to predict the side effects of drugs. Therefore, based on SAO semantic analysis, this paper aims to construct an effective method for identifying potential side effects of drugs. The technical roadmap of the article is shown in Figure 1.

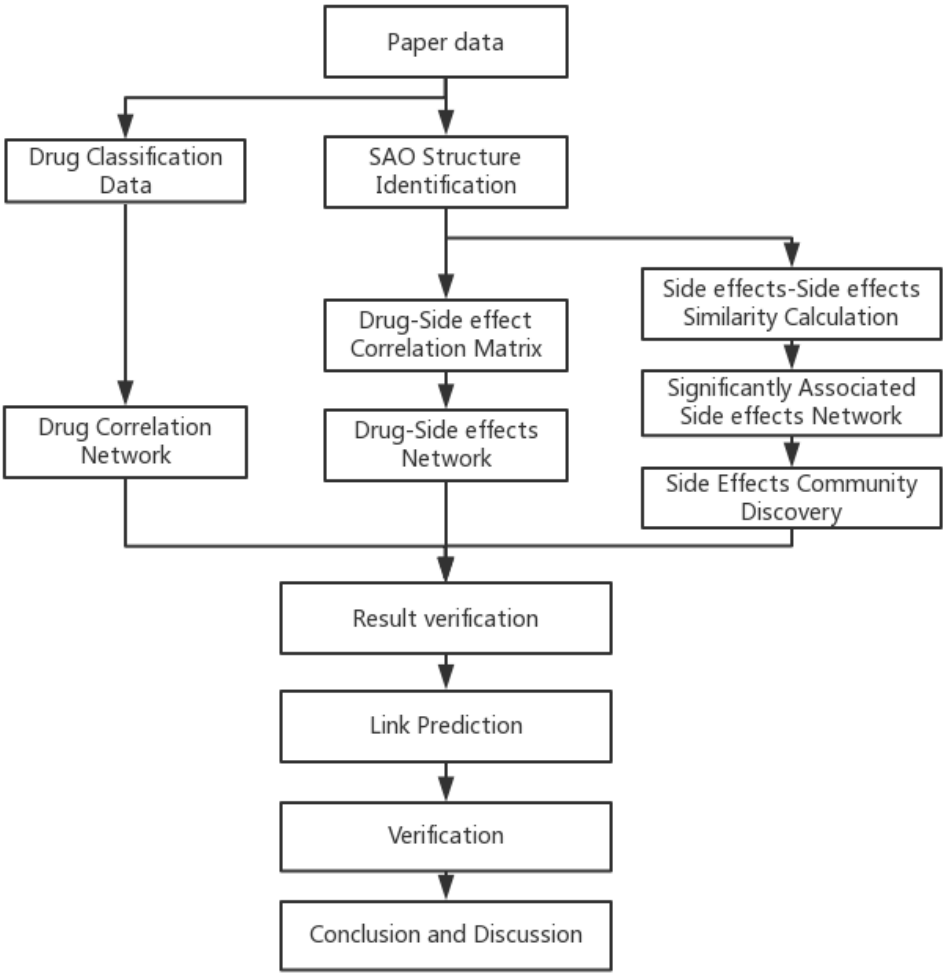

Fig. 1. Technical roadmap

METHODOLOGY

In this paper, we used PubMed data to retrieve all articles related to human drug treatment and side effects during the ten years from 2008 to 2018.

First, the SAO structure was extracted from the biomedical literature using the SemRep tool. Based on the UMLS corpus, combined with the side effect vocabulary in SIDER, the extracted SAO structure was cleaned and screened in order to find the SAO structure related to the side effects of drugs and drugs. A complex network is then established for the "drug-drug", "drug-side effects", and "side effects-side effects" relationships.

The "drug-drug" network was established based on the medicinal chemistry information in DurgBank. According to the extracted SAO semantic structure and the co-occurrence relationship based on "S" and "O", the "drug-side" association matrix is established first, and then the network is established. Inspired by the theory of mapping from binary network to unary undirected graph, we map the "drug-side effects" relationship to the "side effects-side effects" relationship[2] (Figure 2). Based on this network, we measure the cosine similarity between side effects[3], set the relevant

thresholds, and extract a significant "side effect-side effect " network. To detect the community of significantly associated networks of "side effects-side effects", we utilize the edge-cut method. We cluster the related side effects into different sub-networks according to their degrees of association.

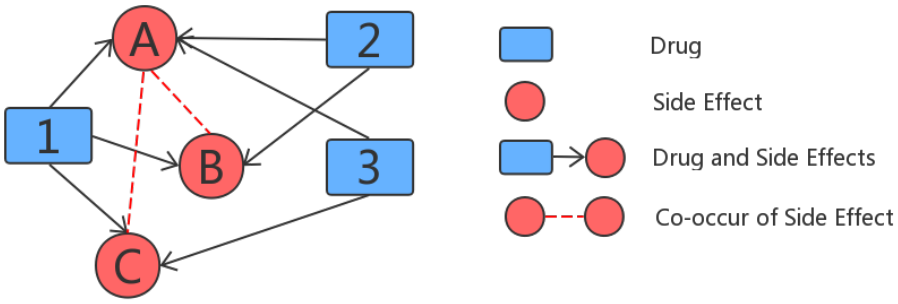

Fig. 2. Side effect correlation diagram

Establish a multi-layer complex network for the above three networks. Link prediction is used to predict pairs of newly created nodes between drugs and side effects. The results of link prediction are evaluated using three indicators: AUC, accuracy, and recall. Compare the predicted Drug and Side Effects with the SIDER database. If it is indeed a potential drug side effect, conduct further literature research to determine whether the prediction is realistic.

Finally, the medical literature is searched for the predicted results to explain and analyze the domain knowledge.

RESULTS AND POLICY IMPACT/IMPLICATIONS

This study can be used to identify drug side effects that are not documented in existing drug-side effects databases, to provide a reference for early detection of potential side effects of drugs, and to provide a viable solution for further use of machine learning to more accurately extract drug-side effects.

REFERENCE

[1] Karimi S, Chen W, Metke-Jimenez A, et al. Text and Data Mining Techniques in Adverse Drug Reaction Detection[J]. Acm Computing Surveys, 2015, 47(4): 1-39.

[2] Luo Y F, Liu Q J, Wu W J, et al. Predicting Drug Side Effects Based on Link Prediction in Bipartite Network[M]. New York: Ieee, 2014: 729-733.

[3] Qabaja A, Jarada T, Elsheikh A, et al. Prediction of gene-based drug indications using compendia of public gene expression data and PubMed abstracts[J]. Journal of Bioinformatics & Computational Biology, 2014, 12(03): 96ra77-.