

2016 International Conference on Computational Science

Identifying Disease -Treatment Relations using Machine Learning Approach

#1Keerrthega.M.C*#2Ms.D.Thenmozhi

#1Student,SSN College of Engineering,Chennai-603110

#2Assistant Professor,SSN College of Engineering,Chennai-603110

Abstract

Identifying the disease treatment relation enables to find what disease a person suffers from and what appropriate treatment can be given to that person. The semantic relation tags namely Cure, Prevent and Sideeffects helps to find out the relationship between disease and treatment. Many methodologies like co-occurrence analysis, rule based methodologies and statistical methods are used in disease treatment relation. However, machine learning is widely used in many applications like protein-protein interaction , extraction of medical knowledge and in health care field. we propose a machine learning approach termed as SMO classification, which uses several features namely medical papers, medical abstracts. Our approach identifies the features namely disease-treat, cure, prevent and sideeffects. The performance can be measured by Accuracy, Precision, F-measure and Recall.

Keywords:Natural Language Processing;Machine Learning;SMO Classification

1. Introduction

Identification of the disease-treatment relation enables to identify what type of disease - treatment relation exists in sentences. The identified disease - treatment relation contains the three semantic relations namely cure, prevent and sideeffects. Disease-treatment relation identification can be used by the public to know about the above types of relations between disease and treatment. Identification of disease - treatment relations is useful to health care providers, private clinics, hospitals, doctors and common people.

2. Literature Survey

A. Semantic Relations in Bioscience text

In this paper [1] mapping biomedical information into structural representation using machine learning

for information extraction. The extraction of medical abstract is obtained through text classification. Semantic lexicons of words labelled with semantic classes where associations can be drawn between words which helps in extracting the necessary sentences related to the query. Naive Bayes (NB) algorithm to extract semantic relation like Gene-Protein from Medline abstracts.

B. Learning to extract relations from Medline

In this paper [5] individual sentences are considered as features processed by the Naive Bayes classifier. Here each feature is considered as positive training set. Extraction of words from Medline abstract has been done by using Naive Bayes, CNB algorithm (Compliment Naive Bayes Classification). It used bag of words during classification but not used natural language processing due to this performance of output degrades.

C. Extraction of Disease-Treatment relations from Biomedical Sentences

In this paper, the dataset is annotated with 8 semantic relations between diseases and treatments, using Hidden Markov models and maximum entropy models to perform both the entity recognition and relation identification. The representation techniques are done by Parts Of Speech (POS), Phrases and terms from MeSH (Medical Subject Headings.)

D. Biomedical Language Processing: What's beyond Pubmed

In this paper [12] It involves natural language processing for processing of biomedical words and in this work it takes the name of disease and give the solution which has been stored in database of that disease by parsing user statement using natural language processing but it does not do diagnosis of disease.

E. Hybrid Machine Learning Implementation for classifying Disease-Treatment relations in Short texts

In this paper, selection techniques are used in order to identify the most suitable words as features by sentence selection and relation identification. The task of relation extraction is tackled in the medical literature focus on biomedical tasks such as sub cellular location, gene-disorder association, diseases and drugs. The data sets used in biomedical specific tasks use short texts.

3. Proposed System

Identifies Disease-Treatment Relation and classifies what type of disease-treatment relation namely cure, prevent and side effects exists in a sentence from Biotextc abstracts using Supervised Approach. The architecture of the proposed system:

Informative- information about the disease or treatment

Non-informative- no information about the disease or treatment.

Semantic Relations: Cure, Prevent, Side Effects.

The modules of the proposed system are described below:

(i) Sentence Selection

(ii) Feature extraction

- BOW Extraction
- Syntactic Representation
- Metamap Representation

(iii) Classification

Download English Version:

<https://daneshyari.com/en/article/485233>

Download Persian Version:

<https://daneshyari.com/article/485233>

[Daneshyari.com](https://daneshyari.com)