

Biomedical Text Mining and Applications in Drug Discovery

Dr. N. Jeyakumar,

Dept. of Bioinformatics, School of Life Sciences, Bharathiar University,
Coimbatore 641 046

Email: n.jeyakumar@yahoo.co.in



Abstract

Besides experimental data, there is a substantial amount of biomedical knowledge recorded only in the form of free-text in abstracts, full-text articles and clinical records etc. For example the biomedical literature database PubMed currently contains about 30 million abstracts and an average about 40,000 to 50,000 new abstracts are added every month. Storage of articles in PubMed has tremendously increasing every year . Biomedical literature data mining is one of the recent challenges in Big Data Analysis. The discipline text mining is evolved for automatic extraction new knowledge from published literature. Text mining is defined as the utilization of automated methods for the enormous amount of knowledge available in text documents. The application of biomedical text mining includes biomedical named entity tagging (e.g. genes, proteins, enzymes, drugs etc.), entity concept relation extraction (e.g. protein-protein interactions, drug-disease relations etc.) and mining new knowledge (e.g. drug targets, biomedical pathways, functions). This presentation serves as an introduction to the applications of various text mining approaches in drug discovery. It is divided into two parts with the first half as an overview of text mining in the biosciences. The second part reviews strategies and methods for few unique applications of text mining in drug discovery.