International Journal of Engineering & Technology, 7 (4.5) (2018) 171-174

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

A Tool for Suggesting Ayurvedic Remedies from Curated and Classified Clinical Trial Reports

Sariga Raj[1]*, Priya P[2]

[1,2]*Division of Information Technology, School of Engineering, Cochin University of Science & Technology, Kochi, Kerala, India*
**Corresponding author E-mail:sariga.raj@gmail.com*

Abstract

It requires great effort to search through huge number of published articles that provide information we need. Therefore it is necessary to find a solution that helps researchers in gaining accurate and deep understanding about diseases. Thus drug discovery and drug repurposing are gaining significance with the current onics tools. Traditional Medical practices like Ayurveda needs to be more visible to practitioners with evidence based approach. The clinical trials conducted have to be shared with the world for attaining the very philosophy of Ayurveda.. This paper presents a survey on various text mining technologies developed to classify theories and literature pertaining to the clinical observations of practitioners and suggests a possible solution to match a patient's symptoms.

Keywords: *Drug discovery; MeSH based Text mining method; Network pharmacology; Text Mining*

1. Introduction

Research in bioinformatics concerns the elucidation of new information management techniques that can aid in the basic life sciences. A major theme of investigation involves extracting and encoding the knowledge of clinical-practice guidelines within computers and the development of information technology that can communicate these "best practices" directly to clinicians at the point of care. Researchers are seeking new ways to represent and archive experimental data so that intelligent software systems can interpret the results and compare experimental outcomes automatically.

With the rapid growth of available biomedical data in the post-genomic era, systems biology and poly-pharmacology have provided fresh insight into the drug discovery [1, 2]. The computational biology provides profitable approach to address the scientific suspense through efficacious modelling and theoretical exploration. In 2007, Hopkins [3] created a novel concept of network pharmacology, which is built on the fundamental concept that many effective drugs in therapeutic areas act on multiple rather than single targets. Network pharmacology can be reconstructed with molecular networks that integrate multidisciplinary concepts including biochemical, bioinformatics, and systems biology [3]. It affords a rewarding assistance to forecast the off-target effects at a higher efficiency, which could improve the potency for drug discovery through a novel network mode of "multiple targets, multiple effects and complex diseases"

In India there is a very strong acceptance of the Traditional Medical System and Practices. Most of these have not been acceptable to Modern Medical practitioners earlier but is soon gaining popularity. The refusal to accept therapeutic values of Traditional Medical system is the lack of scientific evidence. Even though India has a rich heritage and history of Medical Practice very few vaidyas have ventured into curating their clinical trials and experience.[4] There has been a lot of effort to scientifically validate the traditional knowledge by eminent scientists. In their paper [5], Chandran et al., explored the therapeutic properties of Thriphala and established the anti-cancer properties using Network Pharmacological approach. Several similar researches have been conducted in Traditional Chinese Medicines. This paper tries to bring out a design of a software tool that will help in identifying alternate actions and formulations of Ayurvedic medicines based on Complex Networks developed through Text Mining of Ayurvedic Clinical Trials. The proposed system is not expected to perform as a diagnostic tool. The objective of the tool is to search for ayurvedic remedies for symptoms or indications given. The results are obtained on the basis of the clinical trial reports and literature published. A classification of the results are made using MeSH and stored in a database.

The paper explores the state of the art in the next section. The model is discussed in the third section, it's implementation, results and conclusions in the remaining sections of the paper.

2. Literature Survey

Over the past decades, drug discovery has followed the dominant paradigm of the "one gene, one drug, one disease" and mainly focused on designing exquisitely selective ligands which could avoid side effects [1]. [6] reviews the recent boom in network methods helping hit identification, lead selection optimizing drug efficacy, as well as minimizing side-effects and drug toxicity. Successful network-based drug development strategies are shown through the examples of infections, cancer, metabolic diseases, neurodegenerative diseases and aging. As a summarization of greater than 1200 references, it also suggests an optimized protocol of network-aided drug development, and provides a list of systems-level hallmarks of drug quality.

Network pharmacology has emerged as a new topic of study in recent years. It aims to study the myriad relationships among proteins, drugs, and disease phenotypes. Numerous research propose

to establish comprehensive knowledge links between proteins, targets, disease, symptoms, pathways, molecules [7]. Efficiency and efficacy of drugs have been analysed using this approach.[8]

Several text mining services have been developed that helps biologists and researchers search through largevolume of text data. Guang Zheng et al. [9] constructed networks associated with Chinese herbal medicine Danggui,using data slicing algorithm which is based on the principle of co-occurrence. In this paper the authors explored the association rules of Danggui. Text mining is used to demonstrate these rules in different networks. Said Bleik et al. [10] represent a document as a graph that identify high level concepts and build concept graph that contains rich representation of documents. Text categorization is then done using graph kernel techniques. The results show that accuracy improved significantly compared to the techniques of vector representation based on extracted key entities. Lejun Gong et al. [11] presented a text mining approach to extract breast cancer related information as entities and relationship among them. For entity recognition, a system called BerMiner is developed based on CRFs model. The experimental dataset is annotated and it will be useful for research in the field of breast cancer. Qiyu Jiang et al. [12] identified CRF as a suitable model for named entity recognition. They proposed a system for mining clinical terms from TCM medical records. The paper compares various sequence labeling methods: CRF, HMM, ME and MEMM. The results show that, CRF has the best recognition ability compared to others. Zhenchao Jiang et al. [13] presented a system to train word embeddings using biomedical domain-specific word embedding model. They compare word embeddings with other models using two deep learning systems, i.e., Deep Belief Network (DBN) based DDI Extraction model and Recurrent Neural Network (RNN) based NER model. This model is better than other general-purpose word embedding models. This model improves performance of deep learning systems in biomedical text mining.

Thomas Evangelidis et al. [14] developed a Proteome-wide Off-target Pipeline (POP) that helps the identification of drug off-target and polypharmacology drug design. The process of ligand binding site analysis and the protein-ligand docking are done automatically. POP is a valuable tool for drug discovery that performs screening of protein groups with functional similarity or structural features at a faster rate. Yucel Kocyigit and Huseyin Seker[15] developed a hybrid classification model to identify antibiotic drug targets. Support Vector Machine, Linear Discriminative and Naïve Bayesian classifiers are chosen. The hybrid methods proposed is able to handle highly imbalanced data sets. The model has very high accuracy compared to previous models. Lin Wu et al. [16] developed a graph theoretic algorithm to identify the minimum steering node (MSS) for a given network that can be applied to several biomolecular networks. Biomolecules identified in the MSSs play essential roles in controlling the states of the networks. Arezou Koohi et al. [17] introduced an approach of co-clustering of drug, diseases and genes. This method is based on the multi-view graph similarity of gene nodes. To find highly connected gene modules, this system use gene similarity function directly as a part of clustering optimization. The co-clustering method presented in their work is highly scalable and can also be used for other graph mining cases.

Dries Harnie et al. [18] proposed a system that uses well-known machine learning technique such as Spark-driven approach that allows quick data exploration based on Apache Spark. They are convinced that this will yield better results faster. In another paper Lin Wu et al. [19] investigated the MSSs of biomolecular networks by considering the drug-protein binding information. They explain biomolecular network dynamics using linear time-invariant dynamic model. A minimum cost maximum flow algorithm is developed to identify the MSS with steering node preference. These MSSs are enriched with known drug targets compared to the randomly chosen MSSs. Uma Chandran et al. [20] developed pharmacology networks of Triphala, a commonly used Ayurveda formulation. The interaction of bioactives with molecular targets and their relation with diseases can be well understood by analyzing these networks. Such pharmacology networks could be used in gaining insight into the world of drug design. Catalina O Tudor et al. [21] developed a system to get information about specific gene from biomedical literature. This method automatically identifies key information from the gene's literature.

Literature retrieval in the biomedical domain is greatly facilitated by indexing of articles with MeSH terms. MeSH, which is maintained by the National Library of Medicine, forms taxonomy of the biomedical words. MeSH subheadings may provide additional information for each individual article. Several works were introduced using the controlled MeSH concepts to extract information from literatures.

Michael Krauthammer et al. [22] propose a technique that cross-links MeSH terms with protein interaction data extracted from natural language processor for mining biomolecular literature. The system realizes automated information extraction and their storage in a dedicated knowledge base with the help of MeSH's controlled vocabulary. Antonio J Jimeno-Yepes et al. [23] explores the use of automatically generated summaries of biomedical articles for text indexing and categorization. This study compares different indexing algorithms such as MTI (Medical Text Indexer), individual MTI components and machine learning. They state that tuning of the summarization algorithms based on MeSH indexing performs better. Zhong Huang[24] applied semantic based information retrieval method to mine PubMed abstracts to find genes-disease association. This is an effective method to discover disease related biomarker.

Guangyu Shan et al. [25] presents a MeSH based text mining method to predict new prebiotics. The system use a systematic feature-ranking algorithm which classifies a variety of carbohydrates into different clusters according to their chemical and biological attributes. The system is developed using the concept of MedMeSH summarizer, a text mining algorithm that summarizes a group of genes and describes the functionality of the group. An exhaustive text mining method is used to get related MeSH terms and a list is created using optimal terms and is ranked by an overall relevance score. This method used cross-validation ROC analyses to evaluate the performance of the model. The performance is better than that of random forest method. The results show that this method attained a specificity of 0.876 and a sensitivity of 0.838. Horacio Caniza et al.[26] proposed a method that uses the MeSH to accurately quantify the similarity between diseases at molecular level. This method explores the existing information about diseases that is scattered across various biomedical literatures. The performance of this method is better than the simpler, overlap based methods and it also obtains a high-quality score that characterize disease similarity.

K. Jensen[27] presents a systematized approach to find food-disease association to identify novel bioactive compounds from natural sources. It uses text mining and Naive Bayes classification to assemble the information regarding health benefits of vegetables, fruits and other plants to establish relationship between foods, phytochemicals and human diseases. This will help the repurposing of medicinal plants to other diseases and also access the impact of food on health. H. G. G. Vaka[28] describes a text mining process to extract associations among disease and herbs related to Ayurveda using PubMed as literature resource. Automated Vocabulary Discovery algorithm is used to identify biological objects from extracted abstracts.

3. Proposed System

This paper proposes a system that enables the retrieval of information obtained from the literature by means of text mining,

which provide a way for finding appropriate ayurvedic formulations to cure diseases. We present a MeSH based text mining method using the large PubMed repository. MeSH terms could represent the whole text accurately and we can extract characteristic features from vast biomedical literature using these high-quality widgets.

3.1. System Architecture

The proposed system fundamentally is organised in two modules. The Text Mining Module is vested with the responsibility of extracting information from published articles of PubMed and classifying into categories like disease, symptoms, bioactives, targets, proteins, signalling pathways by MesH standards. These are stored in a store.

There has to be periodical updation of this database because the publications happen every day. From the the practitioner side, the symptoms and patient conditions are sent along as a query to obtain disease suggestions or drug suggestions with a confidence factor. Figure 1 represents the the architecture of the proposed system.

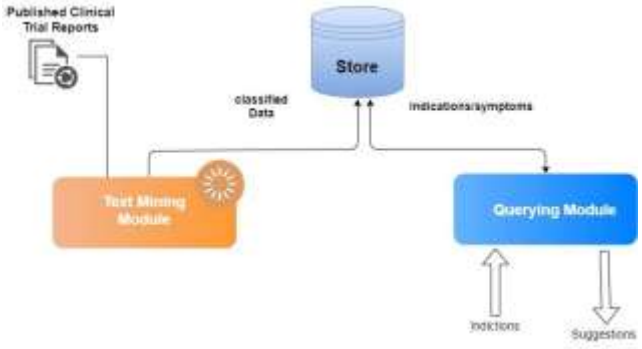

Fig. 1:. System Overview

3.1.1 Text Mining Module

Due to complex naming systems, the biomedical literature is very difficult to explore. However, drug ontology and entity extraction tools are becoming available. Text mining technology and the development of biological and gene ontologies helps classify and query data efficiently. The text mining algorithm identifies association among disease and ayurvedic formulations and also assigns a numerical confidence score to them. Those with higher score is expected to be more effective that that with lower scores. The extracted results can be stored in a database and users can retrieve them effectively through a web interface. This method can drastically reduce the total search time compared to comprehensive search.

First step is to retrieve all articles that contain Ayurvedic terms form PubMed. The number of articles related to ayurveda is very small compared to the total number of articles in PubMed. SVM classifier can be used to discriminate articles on Ayurvedic clinical trials from those of other disciplines. Then download relevant articles and store in a local system and PubMed id‘s are noted. Mining these data will provide association between the terms in extracted data. The results can be stored in a database and user can access the information by querying the required data.

3.1.2 Data Repository

Data repository is created to store the extracted information from retrieved articles. It contains a table to store the citation details. Each record contains the MEDLINE citation PubMed id, year, title, and abstract. Also contains table to store disease-formulation association details. Updation of the records should be done periodically. It is made to communicate with the front end so that the data can be accessed by users.

3.1.3 Query Engine

The query engine provides an interface for the users that help to retrieve required data from the database. Users can query the database for particular indications of disease so that associated formulations can be retrieved. Query engine process user request and results are returned from the store. Based on the indications mentioned, the formulations are ranked to find the most appropriate formulation and the result with highest rank can be selected

3.2. Implementation

The articles are obtained from pubmed using MeSH term search. It is reliable for positive data, as it is verified by NLM. Text mining is performed using Support Vector Machine for classifying the documents and obtain rich data. Datas are stored in database created using MySQL. Retrieved data are loaded to the database and are accessed through a web interface created using python.

3.3. Results

The System Proposed was given 50 clinical trial results obtained from Therapeutic Index of a pharmaceutical which was compiled from the traditional ayurvedic scripts.. As many as 50 drugs or formulations were classified and stored along with the indications or symptoms. When the system was queried with 2-3 indications the result obtained was a suggestion of diseases and medication for the symptoms. The suggestions were ranked according the matching of each indication. Since this was a prototype a weighted average method was used to obtain a rank. Higher weights were assigned to indications, smaller weights for other parameters like age, sex, blood type. Fig. 2 shows the screenshot of the results obtained.

Fig.2: Screenshot of obtained results for the indications fever, cold and cough. Appropriate formulations are listed based on their ranks.

4. Conclusion

The study of the systems biology was the driver to conduct research in the area of pharmacology. Coming from a country like India prompted the authors to study the state of art in bioinformatics related to Ayurveda and other traditional medical knowledge base. A lack of updation of clinical experiments was noticed in most literature. Hence an attempt towards bridging this gap was the mission of the study. As a result a software tool was envisaged to mine the knowledge from emerging publications dedicated to Ayurvedic Research. A Query engine was designed and developed to test the effectiveness of the classification tool. Though the results obtained are not analysed, this paper can be treated as a step forward in this direction.

References

[1] A. Friboulet and D. Thomas, "Systems biology—an interdisciplinary approach", *Biosensors and Bioelectronics*, vol. 20, no. 12, pp. 2404–2407, 2005.

[2] J. T. Metz and P. J. Hajduk," Rational approaches to targeted polypharmacology: creating and navigating protein ligand interaction networks," *Current Opinion in Chemical Biology*, vol. 14, no. 4, pp. 498–504, 2010.

[3] A. L. Hopkins, "Network pharmacology," Nature Biotechnology, vol. 25, no. 10, pp. 1110–1111, 2007.

[4] Bhushan Patwardhan, :Bridging Ayurveda with evidence-based scientific approaches in medicine", The EPMA Journal 2014, 5:19

[5] U Chandran, N Mehendale, G Tillu, B Patwardhan, "Network Pharmacology of Ayurveda Formulation Triphala with Special Reference to Anti-Cancer Property", Combinatorial Chemistry & High Throughput screening, 2015

[6] Peter Csermely, Tamás Korcsmáros, Huba J.M. Kiss, Gábor London, Nussinov," Structure and dynamics of molecular networks: A novel paradigm of drug discovery: A comprehensive review",*Pharmacol Ther.* 2013 June ; 138(3): 333–408. doi:10.1016/j.pharmthera.2013.01.016.

[7] Nir Yosef, Lior Ungar, Einat Zalckvar, Adi Kimchi, Martin Kupiec, Eytan Ruppin and Roded Sharan "Toward accurate reconstruction of functional protein networks" Molecular Systems Biology 5:248; doi:10.1038/msb.2009.3

[8] Emre Guney, Jo¨rg Menche, Marc Vidal and Albert-La´szlo´ Bara´basi "Network-based in silico drug efficacy screening" Nat. Commun. 2016; 7:10331–10343.

[9] G. Zheng, J. Zhan, H. Guo, M. Jiang, C. Lu, and A. Lu, "Exploring associated rules of Danggui in traditional Chinese medicine through text mining," Proc. IEEE Int. Conf. Softw. Eng. Serv. Sci. ICSESS, pp. 198–203, 2013.

[10] . S. Bleik, M. Song, A. Smalter, J. Huan, and G. Lushington, "CGM: A biomedical text categorization approach using concept graph mining," *2009 IEEE Int. Conf. Bioinforma. Biomed. Work.*, pp. 38–43, 2009.

[11] . L. Gong, R. Yan, Q. Liu, H. Yang, G. Yang, and K. Jiang, "Extraction of biomedical information related to breast cancer using text mining," *2016 12th Int. Conf. Nat. Comput. Fuzzy Syst. Knowl. Discov. ICNC-FSKD 2016*, pp. 801–805, 2016.

[12] . Q. Jiang, H. Li, and J. Liang, "Free text mining of TCM medical records based on conditional random fields," pp. 0–5, 2016.

[13] . Z. Jiang, L. Li, D. Huang, and L. Jin, "2015- Training_word_embeddings_for_deep_learning_in_biomedical_text_mining_tasks-1," pp. 625–628, 2015.

[14] T. Evangelidis and L. Xie, "An integrated workflow for proteome-wide off-target identification and polypharmacology drug design," *Tsinghua Sci. Technol.*, vol. 19, no. 3, pp. 275–284, 2014.

[15] Y. Kocyigit and H. Seker, "Hybrid imbalanced data classifier models for computational discovery of antibiotic drug targets," *Conf. Proc. ... Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf.*, vol. 2014, pp. 812–815, 2014.

[16] L. Wu, Y. Shen, M. Li, and F. X. Wu, "Network Output Controllability-Based Method for Drug Target Identification," *IEEE Trans. Nanobioscience*, vol. 14, no. 2, pp. 184–191, 2015.

[17] A. Koohi, H. Homayoun, J. Xu, and M. Orooji, "Co-clustering of diseases, genes, and drugs for identification of their related gene modules," *Proc. 8th Int. Conf. Adv. Comput. Intell. ICACI 2016*, pp. 407–411, 2016.

[18] D. Harnie *et al.*, "Scaling machine learning for target prediction in drug discovery using Apache Spark," *Futur. Gener. Comput. Syst.*, vol. 67, pp. 409–417, 2017.

[19] . L. Wu, L. Tang, M. Li, J. Wang, and F.-X. Wu, "Biomolecular Network Controllability With Drug Binding Information.," *IEEE Trans. Nanobioscience*, vol. 16, no. 5, pp. 326–332, 2017.

[20] U. Chandran, N. Mehendale, G. Tillu, and B. Patwardhan, "Network pharmacology of ayurveda formulation triphala with special reference to anti-cancer property," *Comb. Chem. High Throughput Screen.*, vol. 18, no. 9, pp. 846–854, 2015.

[21] C. O. Tudor, K. Vijay-Shanker, and C. J. Schmidt, "Mining gene-related information from biomedical literature," *Proc. - 2009 IEEE Int. Conf. Bioinforma. Biomed. Work. BIBMW 2009*, p. 342, 2009.

[22] M. Krauthammer, M. D. Pauline, K. Phd, and C. Friedman, "Linking protein interaction data to the MESH hierarchy," p. 5027, 2001.

[23] A. J. Jimeno-Yepes, L. Plaza, J. G. Mork, A. R. Aronson, and A. Díaz, "MeSH indexing based on automatically generated summaries.," *BMC Bioinformatics*, vol. 14, p. 208, 2013.

[24] Z. Huang, "Mining disease associated biomarker networks from PubMed," *Int. Conf. Syst. Biol. ISB*, pp. 15–18, 2013.

[25] G. Shan, Y. Lu, B. Min, W. Qu, and C. Zhang, "A MeSH-based text mining method for identifying novel prebiotics.," *Medicine (Baltimore).*, vol. 95, no. 49, p. e5585, 2016.

[26] . H. Caniza, A. E. Romero, and A. Paccanaro, "A network medicine approach to quantify distance between hereditary disease modules on the interactome," *Sci. Rep.*, vol. 5, no. 1, p. 17658, 2016.

[27] K. Jensen, G. Panagiotou, and I. Kouskoumvekaki, "Integrated Text Mining and Chemoinformatics Analysis Associates Diet to Health Benefit at Molecular Level," PLoS Comput. Biol., vol. 10, no. 1, 2014.

[28] H. G. G. Vaka and S. Mukhopadhyay, "Hypotheses generation pertaining to ayurveda using automated vocabulary generation and transitive text mining," *NBiS 2009 - 12th Int. Conf. Network-Based Inf. Syst.*, pp. 200–205, 2009