

Discovering Complex Relationships of Drugs over Distributed Knowledgebases

Juan Li, North Dakota State University, Fargo, ND, USA

Ranjana Sharma, North Dakota State University, Fargo, ND, USA

Yan Bai, University of Washington Tacoma, Tacoma, WA, USA

ABSTRACT

Drug discovery is a lengthy, expensive and difficult process. Identifying and understanding the hidden relationships among drugs, genes, proteins, and diseases will expedite the process of drug discovery. In this paper, we propose an effective methodology to discover drug-related semantic relationships over large-scale distributed web data in medicine, pharmacology and biotechnology. By utilizing semantic web and distributed system technologies, we developed a novel hierarchical knowledge abstraction and an efficient relation discovery protocol. Our approach effectively facilitates the realization of the full potential of harnessing the collective power and utilization of the drug-related knowledge scattered over the Internet.

Keywords: Distributed Systems, Drug Discovery, Drug Repositioning, Relation, Semantic Web

INTRODUCTION

Drug discovery is a process of discovering and designing drugs. It is generally related to the fields of medicine, pharmacology and biotechnology. Despite the advances in chemical synthesis techniques as well as combinatorial and cheminformatics and understanding of biological systems, drug discovery is still a lengthy, expensive, and difficult process with a low rate of new therapeutic discovery. The process of drug development consists of drug compounds proposal, pre-clinical, clinical trial, and FDA

review and approval (Anyanwu & Sheth, 2003). Usually, 5000 to 10,000 compounds are proposed for a potential drug (DiMasi et al., 2003; DiMasi et al., 1991; Masia 2009; Collier, 2009). An extensive research of the proposed compounds is conducted to select 2.5% to 5% for preclinical trials to be tested on animals. Among the selected compounds for preclinical trial, only 2% may get approved for clinical trial. Finally, only 1 compound becomes an approved drug for the treatment of diseases and use on humans (Collier, 2009). Studies also show that it takes about 15 years from compound proposal

DOI: 10.4018/ijdst.2014010102

to FDA approval of a new drug to treat a disease, and the total cost is between 0.8 to 1 billion US dollars (DiMasi, 2003).

To overcome the aforementioned problems of drug discovery, drug repositioning (Sleigh & Barton, 2010) has been proposed. Drug repositioning is the application of the existing drugs to new indications or new diseases. An existing drug has passed significant pre-clinical and clinical tests. Its toxicity and other effects are already known. Hence, the cost of using it for some other diseases will be much less as compared to developing a drug from scratch (Hopkins, 2008; Boran & Iyengar, 2010; Druker et al., 1996). Conventional drug design follows the principle of “one gene, one disease, one drug”. One drug is targeted for the treatment of one disease caused by one gene (Sleigh & Barton 2010). Unlike conventional drug design, drug repositioning studies interactions of drugs with multiple targets (Hopkins, 2008; Boran & Iyengar, 2010). One drug can be used with multiple diseases. Some repositioned drugs have been approved with new uses that are different from original uses (Chong & Sullivan, 2007; Verma et al., 2005). Here are some examples. Thalidomide normally as a drug for sedation, nausea and insomnia is being used in the treatment of multiple myeloma (Durk, 2006). Acetylsalicylic acid, as a drug for reducing aches and pains and fever is being used in cardiology to prevent heart attacks, strokes and blood clotting due to its antiplatelet activity (Krumholz et al., 1995). Also Miltefosine for the treatment of Cancer has the new indication as Visceral leishmania (Sundar et al., 1998). Another very common example is of Sildenafil which was earlier used for hypertension is now being used for male erectile dysfunction in the name of Viagra (Boolell et al., 1996).

Presently, most of repositioned drugs are developed by observing the side-effects of other drugs (Aronson, 2007; Ashburn & Thor, 2004). However, it will be more effective to study the polypharmacological action of drugs and examine proteins, genes, pathway and other important factors, rather than discovering the effects merely by observation. The hidden

relationship between drugs and diseases could not be observed and identified incipiently, so does the information about the drug and its new applications.

With the development of Semantic Web technologies, more and more Semantic Web data including data related to proteins, genes, drugs, disease are generated. For example, the Bio2RDF project generates a network of the life science data. Different databases across life sciences platform have been linked using open-source Semantic Web technologies to provide support biological knowledge discovery (Nolin et al., 2008). The Linking Open Drug Data project brings the data sources about drugs, Chinese medicine, clinical trials, diseases and pharmaceutical companies together onto the Web of Linked Data and facilitates the integration of about 8.4 million data (Samwald et al., 2011). To effectively utilize the large amount of semantic data, efficient search mechanisms for Semantic Web data have been proposed for both humans and software agents. For instance, the Semantic Search scans objects to capture instances in a given data set (Guha et al., 2003). By utilizing keywords, the Swoogle search engine retrieves semantic entities as Uniform Resource Identifiers (URI) (Ding et al., 2004). To support complex queries over Resource Description Framework (RDF) bases, query languages, such as SPARQL, have been used to express various restrictions on semantic entities and relationships (<http://www.w3.org/TR/rdf-sparql-query/>). These technologies effectively assist human and software agents to locate desirable information from large amounts of semantic data on the Web. However, they may not reveal hidden knowledge large semantic datasets. For instance, knowledge of determining the complex relationships between multiple semantic entities, i.e., “how drugs X and Y are related.”

In Semantic Web, a sequence of complex relationship between the semantic entities forms a semantic association. The entities could be from disparate sources. More than one path may exist between the entities or they may be associated in more than one way. An entity may

16 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/article/discovering-complex-relationships-of-drugs-over-distributed-knowledgebases/104762

Related Content

Energy Efficient Resource Allocation During Initial Mapping of Virtual Machines to Servers in Cloud Datacenters

Nimisha Patel and Hiren Patel (2018). *International Journal of Distributed Systems and Technologies* (pp. 39-54).

www.irma-international.org/article/energy-efficient-resource-allocation-during-initial-mapping-of-virtual-machines-to-servers-in-cloud-datacenters/196266

Design and Implementation of Hybrid Time (HT) Group Communication Protocol for Homogeneous Broadcast Groups

Isamu Tsuneizumi, Ailixier Aikebaier, Makoto Ikeda, Tomoya Enokido and Makoto Takizawa (2011). *International Journal of Distributed Systems and Technologies* (pp. 37-48).

www.irma-international.org/article/design-implementation-hybrid-time-group/55420

Performance Evaluation of Full Diversity QOSTBC MIMO Systems with Multiple Receive Antenna

Hardip K. Shah, Tejal N. Parmar, Nikhil Kothari and K. S. Dasgupta (2011). *International Journal of Grid and High Performance Computing* (pp. 29-38).

www.irma-international.org/article/performance-evaluation-full-diversity-qostbc/60253

Modeling Scalable Grid Information Services with Colored Petri Nets

Vijay Sahota, Maozhen Li and Marios Hadjinicolaou (2012). *Grid and Cloud Computing: Concepts, Methodologies, Tools and Applications* (pp. 701-716).

www.irma-international.org/chapter/modeling-scalable-grid-information-services/64510

Model-Based Application Deployment on Cloud Computing

Aouat Asmaa, Deba El Abbassia, Benyamina Abou EL Hassan and Benhamamouch Djilali (2019). *International Journal of Distributed Systems and Technologies* (pp. 110-127).

www.irma-international.org/article/model-based-application-deployment-on-cloud-computing/226964