Link Prediction-Based Synergistic Drug Combination Recommendation for the Diagnosis and Treatment of Disease Complications: A Postprint

Authors: Lei Ming, Xia Mengge, Wang Xuefeng, Liu Jia

Date: 2023-04-01T16:02:52+00:00

Abstract

[Purpose/Significance] Drug combinations offer multiple advantages over single-agent therapy in clinical practice; however, the rapid growth in the number of drugs poses challenges for drug combination screening and recommendation. Therefore, designing effective predictive methods to recommend drug combinations with higher synergistic potential for drug discovery researchers is of significant importance for improving screening efficiency. [Method/Process] To address the clinical needs of treating disease complications, we construct a synergistic drug combination recommendation model based on link prediction. First, SAO semantic mining is employed to identify complication information from medical literature. On this basis, a "disease-drug-target" heterogeneous network is constructed using medical databases, and link prediction methods are introduced to evaluate the similarity of drug action mechanisms within the network, predicting which drug combinations are more likely to produce synergistic effects. Subsequently, drug combination recommendations are made for specific diseases or pairs of complications based on the prediction results. [Result/Conclusion] Empirical analysis results from intestinal disease data demonstrate that the synergistic drug combination prediction model is effective and practical.

Full Text

Preamble

Volume 65, Issue 12, June 2021
ChinaXiv Cooperative Journal

Research on Collaborative Drug Combination Recommendation Based on Link Prediction: For Disease Complication Diagnosis and Treatment

Lei Ming[1], Xia Mengge[1], Wang Xuefeng[1], Liu Jia[2]
[1] School of Management and Economics, Beijing Institute of Technology, Beijing 100081
[2] State Key Laboratory of Media Convergence and Communication, Communication University of China, Beijing 100024

Abstract:
[Purpose/Significance] Drug combinations offer multiple advantages over single drugs in clinical treatment, but the rapid growth in drug numbers poses challenges for drug combination screening and recommendation. Therefore, designing effective prediction methods to recommend drug combinations that are more likely to produce synergistic effects for pharmaceutical researchers is of great significance for improving screening efficiency. [Method/Process] To address the needs of disease complication diagnosis and treatment, we constructed a synergistic drug combination recommendation model based on link prediction. First, SAO semantic mining was used to identify complication information in medical literature. On this basis, medical databases were utilized to construct a "disease-drug-target" heterogeneous network, and link prediction methods were introduced to evaluate the similarity of drug action mechanisms in the network, predicting which drug combinations are more likely to produce synergistic effects. Then, based on the prediction results, drug combinations were recommended for specific diseases or pairs of complications. [Result/Conclusion] Empirical analysis of intestinal disease data verified the effectiveness and practicality of the synergistic drug combination prediction model.

Keywords: SAO semantic mining, link prediction, social network analysis, drug combination recommendation, disease complications
Classification Number: G250
DOI: 10.13266/j.issn.0252-3116.2021.12.012

This research is supported by the National Natural Science Foundation of China General Program "Research on Identification Methods of Potential Disruptive Technologies in Biomedical Field" (Project No.: 72074020).

Author Contributions:
Lei Ming: Wrote and revised the paper;
Xia Mengge: Conducted case studies, data analysis, and paper writing and revision;
Wang Xuefeng: Proposed the research topic and writing ideas, finalized the paper;
Liu Jia: Organized literature and revised the paper.

2 Method Framework

The model construction process is illustrated in [Figure 1: see original paper]. The methodology consists of four main components: data acquisition and preprocessing, SAO structure extraction and complication network construction, disease-drug-target heterogeneous network construction, and synergistic drug combination recommendation.

2.1 Data Acquisition and Preprocessing

We retrieved medical literature from the PubMed database and utilized the Comparative Toxicogenomics Database (CTD) and DrugBank for network construction. The data acquisition process involved using Java to implement batch retrieval and preprocessing. The initial dataset was filtered to remove duplicates and irrelevant entries, ensuring data quality for subsequent analysis.

2.2 SAO Structure Extraction and Complication Network Construction

2.2.1 SAO Structure Extraction SAO (Subject-Action-Object) structures represent semantic relationships between biomedical entities. We employed SemRep, a tool that leverages the Unified Medical Language System (UMLS), to extract SAO triples from medical literature. Each SAO structure consists of a subject (S), action (A), and object (O), capturing relationships such as "disease-coexists_{with}-disease." For example, the semantic relationship "Enteritis coexists_{with} Chronic disease of immune system" would be extracted as an SAO triple, enabling the identification of disease complications.

2.2.2 Complication Information Identification The extracted SAO structures were used to build a complication network. Disease nodes were connected if SAO analysis revealed a co-occurrence or causal relationship. The network construction process involved parsing SAO triples, identifying disease-disease relationships, and visualizing the complication network using VOSviewer. This network captures the complex relationships between diseases and their complications, providing a foundation for targeted drug combination recommendations.

2.3 Disease-Drug-Target Heterogeneous Network Construction

We constructed a three-layer heterogeneous network integrating diseases, drugs, and targets. As shown in [Figure 3: see original paper], the network consists of: - Disease nodes connected to drug nodes (indicating treatment relationships) - Drug nodes connected to target nodes (indicating mechanism of action) - Target nodes interconnected based on biological relationships

Data from CTD provided disease-drug associations, while DrugBank supplied drug-target interactions. This heterogeneous network enables comprehensive analysis of drug action mechanisms and their relationships to disease treatment.

2.4 Synergistic Drug Combination Recommendation

2.4.1 Link Prediction Metric Selection We employed multiple link prediction metrics to evaluate the likelihood of synergistic relationships between drug pairs. The metrics and their formulas are summarized in :

Table 1: Link Prediction Metrics and Formulas

Metric	Formula
Common Neighbors (CN)	$CN_{xy} = \lvert\Gamma(x) \cap \Gamma(y)\rvert$
Salton	$\text{Salton}_{xy} = \frac{\lvert\Gamma(x)\cap\Gamma(y)\rvert}{\sqrt{k_x \cdot k_y}}$
Jaccard	$Jaccard_{xy} = \frac{\lvert\Gamma(x)\cap\Gamma(y)\rvert}{\lvert\Gamma(x)\cup\Gamma(y)\rvert}$
Adamic-Adar (AA)	$AA_{xy} = \sum_{z\in\Gamma(x)\cap\Gamma(y)} \frac{1}{\log k_z}$
Preferential Attachment (PA)	$PA_{xy} = k_x \cdot k_y$
Local Path (LP)	$LP_{xy} = (A^2)_{xy} + \alpha(A^3)_{xy}$
Katz	$Katz_{xy} = \sum_{l=1}^{\infty} \beta^l \cdot \lvert paths^l_{xy}\rvert = \beta A + \beta^2 A^2 + \beta^3 A^3 + \cdots$
Local Random Walk (LRW)	$LRW_{xy} = q_x \cdot \pi_{xy}(t) + q_y \cdot \pi_{yx}(t)$
Superposed Random Walk (SRW)	$SRW_{xy} = \sum_{l=1}^{t} LRW_{xy}(l)$

Where $\Gamma(x)$ and $\Gamma(y)$ represent the neighbor sets of nodes x and y, k_x and k_y denote their degrees, A is the adjacency matrix, α and β are tuning parameters, and $\pi_{xy}(t)$ represents the transition probability.

Model performance was evaluated using AUC (Area Under the ROC Curve), which measures the probability that a randomly chosen missing link has a higher prediction score than a randomly chosen non-existent link.

2.4.2 Synergistic Drug Combination Recommendation The recommendation process involves three steps: First, calculate link prediction scores for all possible drug pairs in the heterogeneous network. Second, rank drug combinations based on their scores. Third, recommend top-ranked combinations for specific diseases or complication pairs. As illustrated in [Figure 5: see original paper], for a given disease, the model identifies drug pairs with high synergy potential by analyzing their shared targets and network proximity.

---

3 Empirical Study

We validated our model using intestinal disease data from PubMed, retrieving 42,621 relevant articles published between 1900/01/01 and 2019/07/29. After preprocessing, we extracted 23,365 SAO structures, identifying 14,426 complication relationships. The resulting complication network contained 1,221 disease nodes with 2,336 connections.

3.1 Network Construction

The intestinal disease complication network is visualized in [Figure 6: see original paper], showing complex interconnections between conditions such as Crohn's disease, ulcerative colitis, and irritable bowel syndrome. The disease-drug network ([Figure 7: see original paper]) comprised 153 diseases and 1,532 drugs, while the full disease-drug-target heterogeneous network ([Figure 8: see original paper]) integrated 3,654 targets.

3.2 Link Prediction Results

We evaluated nine link prediction metrics on the intestinal disease dataset. As shown in , Katz achieved the highest AUC of 0.956846, followed by Salton (0.923104) and LP (0.919334). Common neighborhood-based metrics (CN, AA) also performed well, while preferential attachment showed lower performance.

Table 2: AUC Values of Link Prediction Metrics

Metric	AUC
Katz	0.956846
Salton	0.923104
LP	0.919334
AA	0.918104
CN	0.483125
Jaccard	0.483122
SRW	0.917472
LRW	0.482945
PA	0.898134

3.3 Drug Combination Recommendations

Based on Katz scores, we identified top synergistic drug combinations for intestinal diseases. For sclerosing cholangitis and pancreatitis, the model recommended combinations such as: - Rofecoxib + Tibolone - Ursodeoxycholic acid + Pancrelipase - Lynerol + Simvastatin

These recommendations align with known therapeutic strategies, demonstrating the model's practical utility. The high AUC values confirm that link prediction effectively identifies drug pairs with synergistic potential by leveraging network topology and mechanism similarity.

The empirical results validate that our approach can efficiently screen drug combinations, reducing experimental costs and accelerating the discovery of effective therapies for complex disease complications.

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv — Machine translation. Verify with original.