Pharmacological Research 103 (2016) 105–113

Contents lists available at ScienceDirect

Pharmacological Research

journal homepage: www.elsevier.com/locate/yphrs

Drug-symptom networking: Linking drug-likeness screening to drug discovery

Xu Xue[a,b,1], Zhang Chao[c,1], Li PiDong[d], Zhang FeiLong[d], Gao Kuo[b], Chen JianXin[b,*], Shang HongCai[a,*]

[a] Key Laboratory of Chinese Internal Medicine of Ministry of Education and Beijing, Dongzhimen Hospital, Beijing University of Chinese Medicine, Beijing 100700, China
[b] Beijing University of Chinese Medicine, Beijing 100029, China
[c] School of Mathematical Sciences, Dalian University of Technology, Dalian, Liaoning 116023, China
[d] College of Life Science, Northwest A&F University, Yangling, Shannxi 712100, China

ARTICLE INFO

Article history:
Received 24 September 2015
Received in revised form 26 October 2015
Accepted 11 November 2015
Available online 23 November 2015

Keywords:
Drug–symptom relationship
Drug-likeness
SVDD classification
Text mining
Herbal molecule

ABSTRACT

Understanding the relationships between drugs and symptoms has broad medical consequences, yet a comprehensive description of the drug-symptom associations is currently lacking. Here, 1441 FDA-approved drugs were collected, and PCA was used to extract 122 descriptors which explained 91% of the variance. Then, a *k*-means++ method was employed to partition the drug dataset into 3 clusters, and 3 corresponding SVDD models (drug-likeness screening models) were constructed with an overall accuracy of up to 95.6%. Furthermore, 6878 herbal molecules from the TcmSP™ database were screened by the above 3 SVDD model to obtain 5309 candidate drug molecules with highly accept classification of 77.19%. To assess the accuracy of the SVDD models, 8559 herbal molecule-symptom co-occurrences were mined from Pubmed abstracts, involving 697 herbal molecules and 314 symptoms. Most of the 697 herbal molecules could be found in the accepted SVDD data (5309 molecules), showing the potential of the SVDD for the screening of drug candidates. Moreover, a herbal molecule-herbal molecule network and a herbal molecule-symptom were constructed. Overall, the results provided a new drug-likeness screening approach independent to abnormal training data, and the comprehensive collection of herbal molecule-symptom associations formed a new data resource for systematic characterization of the symptom-oriented medicines.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

Symptoms, subjective experiences reflecting changes in the biopsychosocial functioning, sensations, or cognition of an individual [1], serve as particularly important guidelines for the diagnosis, evaluation, prevention, and treatment of diseases. Clearly defining symptoms, accurately understanding symptom differences and correctly recognizing the symptom–drug relationships have critical effects on reducing medication-related problems. Without an effective system for symptom analysis, no significant benefit will be achieved; even worse, inappropriate drug use can occur and resources can be unnecessarily wasted.

Substantial research effort has focused on symptoms and symptom-based diagnosis and treatment [2,3]. The development of multidimensional symptom models drives advancement in symptom management research. Associations of symptom dimensions with IFN-α and ribavirin were documented [4]. Differential efficacy of drugs was found based on symptom dimensional measures for escitalopram and nortriptyline [5] and escitalopram and nortriptyline [6]. Nosological analysis was performed for obsessive–compulsive disorder [7], autism spectrum disorder [8] and somatization [9]. In addition, examples of symptom-based clinical practice guidelines involve therapy for cardiovascular disease [10,11], treatment of infectious diseases [12,13], and management of cancer pain [14,15]. The symptom-based research gives an informative picture of the clinical presentation and a quantitative estimate of the severity of the disability, and also proposes novel diagnostic criteria for diseases.

It is fair to say that to date understanding of symptom-drug relationships has only been partly achieved and much remains to be

* Corresponding authors.
E-mail addresses: cjx@bucm.edu.cn (J. Chen), shanghongcai@foxmail.com (H. Shang).
[1] The authors contributed equally to this work.

http://dx.doi.org/10.1016/j.phrs.2015.11.015
1043-6618/© 2015 Elsevier Ltd. All rights reserved.

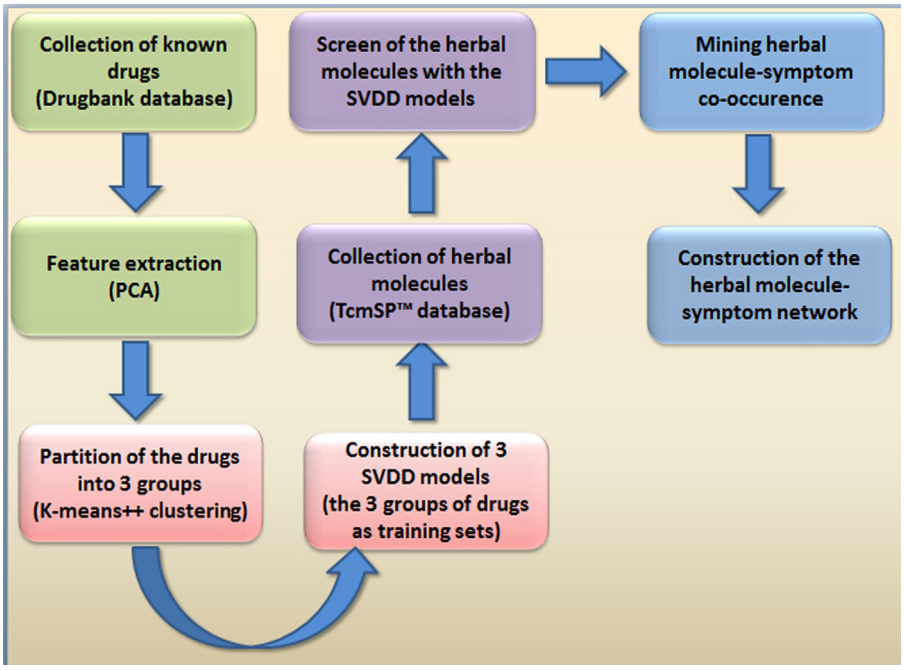

Fig. 1. Schematic representation of the herbal molecule screening workflow. 1441 FDA-approved drugs were first collected from the Drugbank database, and 1441 descriptors were calculated for each drug. To improve the robustness of the data, PCA was used to extract the features, and 122 descriptors were obtained for each drug. Then, the *k*-means++ method was employed to classify the 1441 drugs into 3 groups, and 3 corresponding SVDD models were constructed for the drug groups. Furthermore, 6878 herbal molecules were collected from the TcmSP™ database, and the above 3 SVDD models were used to screen the herbal molecules. 5309 herbal molecules were recognized to be candidate drugs by the SVDD models. To assess the accuracy of the SVDD models, the text-mining method was used to search the relationships of the 5309 herbal molecules with their symptoms, and 697 herbal molecules related to the symptoms were retrieved after quality assessment. To further clarify the herbal molecule-symptom relationships, a herbal molecule-herbal molecule co-occurrence network was constructed based on the symptom similarity of the 697 herbal molecules.

done when it comes to systematic analysis. Owing to the inherent complex interplay between symptoms and drugs, a general understanding of their relationships requires a systems level view rooted in the large-scale information retrieval.

The objective of this work is to lay the ground for a systematic computational approach able to account for such relationships. Typical techniques to investigate biosystems are text mining, information extraction, and retrieval applications, techniques that extract valuable information from large volumes of data [16]. Con-

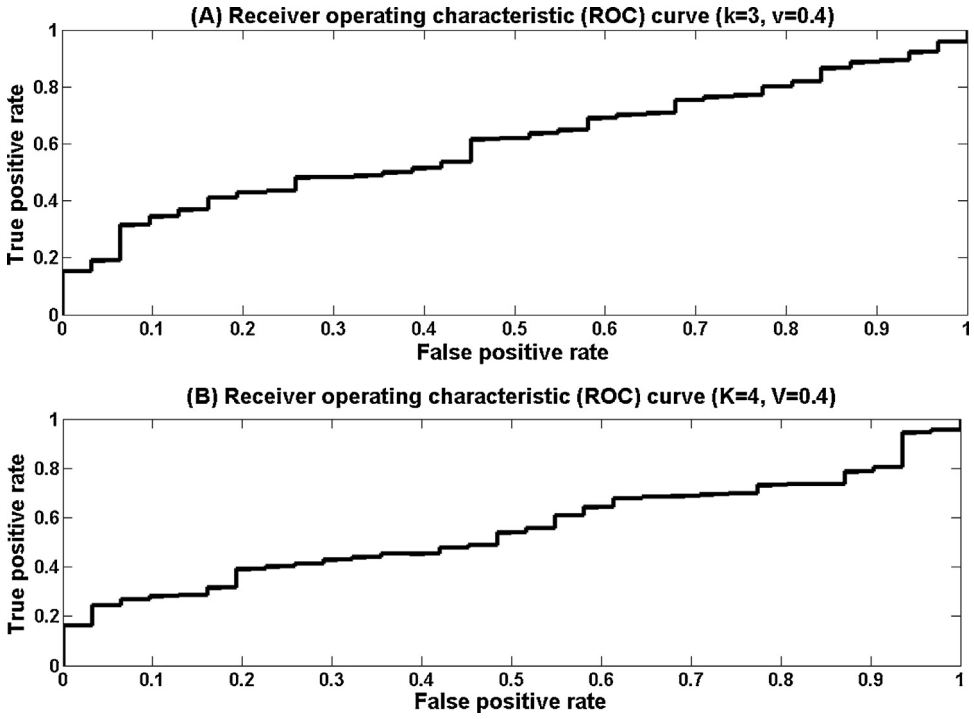

Fig. 2. ROC curves to compare the performance of SVDD. (A) and (B) shows SVDD with $k = 3$ and $v = 0.3$, $k = 4$ and $v = 0.3$, respectively.

Download English Version:

https://daneshyari.com/en/article/5843034

Download Persian Version:

https://daneshyari.com/article/5843034

Daneshyari.com