

The Study on Anomaly Index Mining Research of Coronary Heart Disease with Unstable Angina Pectoris with Qi-stagnation and Blood-stasis Syndrome Based on Co-occurrence Analysis

Xing Zhai ^{1, a}, Youliang Huang ^{1, b}, Zhongzheng Jin ^{1, c}, Fengying Guo ^{1, d, *}

¹Information Center, Beijing University of Chinese Medicine, Beijing, 100029, China

^a zhaix@bucm.edu.cn , ^b huangyl@bucm.edu.cn , ^c jinzz@bcum.edu.cn,

Corresponding author (*): ^d hellofengying@126.com

Keywords: unstable angina pectoris of coronary heart disease; stagnancy of qi and blood-stasis; literature mining; biological index

Abstract. purpose: This research focused on unstable angina pectoris of Qi-stagnation and blood-stasis coronary heart disease, discussed the feasibility of seeking biological markers associated with the syndrome with literature mining methods. Method: using a combination methods of literature research and the clinical research to mine by literature the specific clinical indicators of the stagnancy of qi and blood-stasis in, then compared with the clinical testing of the indicators. Results: We obtained 67 specific index by literature mining and 10 specific index through clinical testing. Of which 8 index that account for results of clinical tests of 80% were mined by two ways. Conclusion: the literature mining laws as the discovery of syndrome of abnormal biological indicators is a new feasible method, but which precision should be improved, it is recommended that further improvement of the mining algorithm, and to extend it to the studies of other diseases and syndromes to validate its' the applicability and reliability.

Introduction

In the biomedical field, text mining's main task is to consolidate and analyze the massive biomedical data and to mine experimental hypothesis and suggestions, in order to verify new scientific discoveries for biologists, so as to raise awareness of biomedical phenomena for people [1]. In 1988, Professor Swanson found the medical relationship between such that as magnesium deficiency and migraine in the biomedical literature in MEDLINE [2], and obtained experimental verification that provides a classic successful precedent for text mining of ground-breaking research. In TCM field, by full mining the mass literature in MEDLINE and in NEI the cold, Shao LI built a interactional hot-cold syndrome network model based on the NEI [3], and proved preliminarily that "disease" and "syndrome" has the feasibility of "disease syndrome combines" on the level of biological network. Thus, powerful text mining technology plays an increasingly important role in modern biomedical research.

This study focuses on the common syndrome of unstable angina pectoris of coronary heart disease, for example, "qi-stagnation and blood-stasis", which searched relevant literature in PubMed to build biological nonspecific clinical index networks of Qi-stagnation and blood-stasis by literature mining methods. On the same time, and Obtained information on cases of unstable angina pectoris of Qi-stagnation and blood-stasis syndrome of coronary heart disease to validate the results of the literature study, thus aims to explore the literature mining methods for accuracy and reliability, and to provide an objective research of syndrome with a new research ideas and methods.

Method

The literature data collection and mining

The Blackboard Learning System is a

1) The construction of qi deficiency literature database

This paper Collected and collated the related terms about "qi-stagnation and blood-stasis syndrome" (on the basis of the criterion of "guiding principle" of clinical research on new drugs of [4]) (Table 1), respectively, took these terms as keywords to search in the PubMed database (<http://www.ncbi.nlm.nih.gov/pubmed>) , and downloaded the article title or abstract containing these key words which are relate with the stagnation of Qi and blood-stasis syndrome for 27691 and are saved with XML file format, then, cleaned and sorted data to construct the relevant literature database of qi-stagnation and blood-stasis. The finish time of literature search and download is in October 11, 2014, so, the data results shall prevail.

Table1. Key words related to “qi-stagnation and blood-stasis syndrome”

Items	Key words
stagnation of qi and blood-stasis	Chest pain / chest tightness, fullness in chest and hypochondrium, Palpitations, Purple tongue,Pulse acerbity

2) literature mining method

This paper took “co-occurrence” as the basic principle of literature mining. “co-occurrence” refers to that these two words are thought as be semantically interrelated if two words often appear together(co-occurrence) in the same window units(such as certain word spacing, a sentence and a document) in the large scale corpus [5]. Based on the above information, this paper argues that an index has likely certain degree associated with qi-stagnation and blood-stasis if it is mentioned several times in the relevant literature, which means that the index may be an abnormal one that leads to Qi-stagnation and blood-stasis. Based on above assumption, using 149 routine clinical index collected during the preliminary epidemiological survey as a probe, we scanned documents about qi stagnancy and blood-stasis syndrome in the document database, and built a biological network by using co-occurrence analysis principles to construct molecular biology network. And literature mining and biological network construction of the main part were carried out in the early development of our literature mining software (PubMed Miner), and literature mining process as shown in Figure 1:

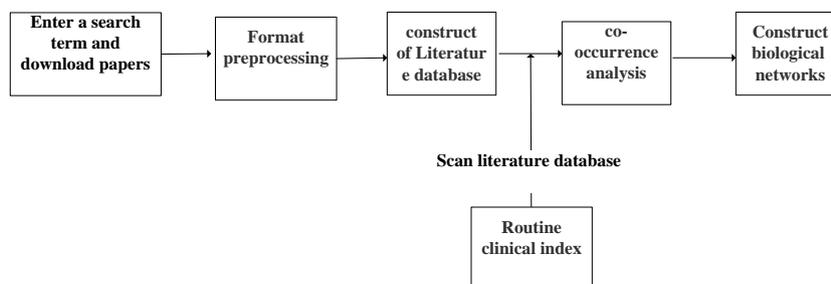


Fig 1 Literature mining flow chart

Clinical data collection and statistics

In the preliminary work, we have collected clinical data of 411 hospitalized patients which are all 20-90 years old, open to sex, with Coronary heart disease and unsteady type angina pectoris standard and voluntary voluntarily signed informed consent.Hospital within 24 hours, 3 Clinician who must have the professional title of associate chief doctor or above would give a judgment for patients about TCM syndrome / TCM syndrome elements and syndrome differentiation of TCM basing on disease diagnosis.

Coronary heart disease and angina pectoris diagnosis criteria refers to criteria as follows: Guidelines for the diagnosis and treatment of chronic stable angina pectoris which were union drawn up by the American college of cardiology(ACC)/American Heart Association(AHA)/ America College of Physicians(ACP)/American Department of Internal Medicine Institute(ASIM).[6]; Diagnosis and treatment recommendations of unstable angina pectoris which were drawn up by Cardiology branch of Chinese Medical Association in 2000[7].

Angina pectoris of coronary heart disease TCM syndrome diagnostic criteria refers to terminology of traditional Chinese medicine clinical diagnosis – syndrome part which were drawn up by Management of National Technical Supervision Bureau in 1997[8]; differentiation of symptoms and signs for classification of syndrome refers to the guiding principle of clinical research for new drug of Traditional Chinese Medicine in treating chest pain (coronary heart disease and angina pectoris) in guiding principle of clinical research on new drugs of Traditional Chinese Medicine which were drawn up in 2002[4].

Result

After artificially screening the index mined by software, we constructed biological index network diagram of the syndrome of qi-stagnation and blood-stasis (Figure 2). From the diagram we drew some conclusions as follows. The syndrome of qi-stagnation and blood-stasis network is made of 67 nodes and 188 sides. Each node represents a clinical routine index mined from literature database. The connection (side) between two indexes represents that two indexes all appeared in the title / abstract in the same article. "Degree" is the most simple and the most important one among the property of complex network node. Value k of a node named "I" is defined as the number of nodes or the number of a side connected with the node. So, intuitively, the greater degree of a node means that the node is more important [9]. In the previous studies [10], we took T-test for routine clinical index of the stagnancy of Qi and blood-stasis and the non-stagnancy of Qi and blood-stasis. The results found 10 index among 149 routine clinical index in the test were statistically significant. Among which, CO2CP, PTA, INR and QT are interval, LVIDS, E peak and A peak have significant differences ($P < 0.05$), BSA%, GGT and CK have significant differences ($P < 0.01$). At last, we reached a conclusion by statistics. Literature mining index is 67 which account for 45% of the total index. Abnormal clinical index of a test is 10 which account for 6.7% of the total index. The index mined by two methods is 8 which account for 12% of Literature mining, for 80% of clinical mining and for 5% of the total index.

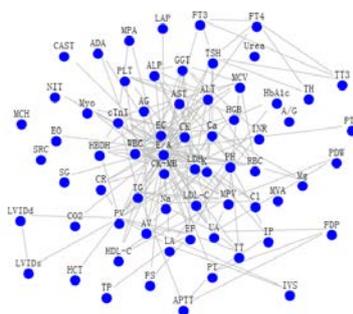


Fig 2 Network diagram of the syndrome of qi-stagnation and blood-stasis

Conclusions

In the past, the traditional method pure rely on clinical verification method for anomaly index of mining, which have some shortcomings as high experimental complexity and the needing huge manpower, material resources and financial. Therefore, we proposes a new method to test the patient anomaly index. That is, we changed the past traditional method that tested all conventional index of patient, Instead of firstly preliminary screened disease index by literature mining to obtain the index that may be associated with the disease or symptom. (After literature mining, more than 50% of the index test can be reduced), and then validating based on these clinical index will can not only reduce the complexity of the experiment, improve the experiment efficiency, and can ensure the accuracy of experimental result. With respect to the clinical verification, it has simple, time-saving, labor-saving features.

But as a new method of mining biological indexes for syndrome, there are some problems in this study. It is mainly reflected two aspects in the deficiency of "depth" and "breadth". In the "depth"

aspect in this test, abnormal index of literature mining contains only 80% abnormal indexes of clinical laboratory, hasn't reached 100%. So the accuracy of literature mining remains to be further improved. Later, we will consider to improve existing literature mining algorithm and expand the search scope of document; in the "breadth" aspect, this research only studied one type of syndrome of qi-stagnation and blood-stasis". In the next step we will expand the research to other syndromes, thus, thus, will further verify the generality and reliability of this method.

Acknowledgements

This work was financially supported by the the Beijing Young Talent Program (YETP0821), the Beijing University of Chinese Medicine "Young Teachers of Special Plan" (QNJSZX016), Beijing University of Chinese Medicine "Young Teachers of Special Plan" (QNJSZX008),

References

- [1] Li S, Zhang Z Q, Wu L J, et al. Understanding ZHENG in traditional Chinese medicine in the context of neuro-endocrine-immune network[J]. *Systems Biology, IET*, 2007, 1(1): 51-60.
- [2] HaoChang WANG, TieJun Zhao. Research and development of biomedical textmining technology [J]. *Journal of Chinese information*, 2008, 22(3): 89-98.
- [3] Smalheiser N R, Swanson D R. Assessing a gap in the biomedical literature-Magnesium-deficiency and neurologic disease[J]. *Neuroscience research communications*, 1994, 15(1): 1-9.
- [4] XiaoYu Zheng. Guiding principle of clinical research on new drugs of traditional Chinese Medicine [M]. Beijing: Chinese pharmaceutical science and Technology Publishing House, 2002: 41.
- [5] Bin Qi, Lv Ting. Co-occurrence analysis and application in data of biomedical information in text mining [J]. *Chinese Journal of Medical Library and Information Science*, 2009 (3): 41-43.
- [6] Gibbons RJ, Chatterjee K, Dale J, et al. ACC/AHA/ACP-ASIM guidelines for the management of patients with chronic stable angina [J]. *J Am Coll Cardiol*, 1999, 33(7): 2092-2197.
- [7] Cardiology branch of Chinese Medical Association. Recommendations about diagnosis and treatment of unstable angina pectoris [J]. *Chinese Journal of cardiovascular disease*, 2000, 28(6): 409-412.
- [8] WenFeng Zhu, YongYan Wang. Terminology of traditional Chinese medicine clinical diagnosis - Part of syndrome [M]. Beijing: China Standards Press, 1997: 17-20.
- [9] Fei Wang. Analysis of the innovation network structure and evolution characteristics of Medicine [J]. *Social Sciences in Nanjing*, 2011(1): 149-155.
- [10] Shi Qi. Traditional Chinese Medicine Syndrome Identification Modes of Unstable Angina Pectoris Caused by Coronary Heart Disease Based on Data Mining[D]. Beijing: Beijing University of Chinese Medicine, 2012