COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE 104 (2011) 286–291

ELSEVIER

journal homepage: www.intl.elsevierhealth.com/journals/cmpb

Cross-domain probabilistic inference in a clinical decision support system: Examples for dermatology and rheumatology

Ying-Jui Chang[a,b,c], *Min-Li Yeh*[a,b,d], *Chyou-Shen Lee*[e,f], *Chien-Yeh Hsu*[b,j], *Yu-Chuan (Jack) Li*[b,g,*], *Wen-Ta Chiu*[h,i]

[a] *Graduate Institute of Medical Science, College of Medicine, Taipei Medical University, Taiwan*
[b] *Graduate Institute of Biomedical Informatics, College of Medical Science and Technology, Taipei Medical University, Taiwan*
[c] *Department of Dermatology, Far Eastern Memorial Hospital, Taiwan*
[d] *Department of Nursing, Oriental Institute of Technology, Taiwan*
[e] *Department of Internal Medicine, Division of Allergy, Immunology and Rheumatology, Mackay Memorial Hospital, Taiwan*
[f] *Mackey Medicine, Nursing and Management College, Taiwan*
[g] *Department of Dermatology, Taipei Medical University, Wan Fang Hospital, Taiwan*
[h] *School of Medicine, Taipei Medical University, Taiwan*
[i] *Graduate Institute of Injury Prevention and Control, Taipei Medical University, Taiwan*
[j] *Center of Excellence for Cancer Research (CECR), Taipei Medical University, Taiwan*

ARTICLE INFO

Article history:
Received 9 April 2011
Received in revised form
14 July 2011
Accepted 14 July 2011

Keywords:
Bayes formulation
Clinical decision support system
Knowledge base

ABSTRACT

Introduction: Maintaining a large diagnostic knowledge base (KB) is a demanding task for any person or organization. Future clinical decision support system (CDSS) may rely on multiple, smaller and more focused KBs developed and maintained at different locations that work together seamlessly. A cross-domain inference tool has great clinical import and utility.
Methods: We developed a modified multi-membership Bayes formulation to facilitate the cross-domain probabilistic inferencing among KBs with overlapping diseases. Two KBs developed for evaluation were non-infectious generalized blistering diseases (GBD) and autoimmune diseases (AID). After the KBs were finalized, they were evaluated separately for validity.
Result: Ten cases from medical journal case reports were used to evaluate this "cross-domain" inference across the two KBs. The resultant non-error rate (NER) was 90%, and the average of probabilities assigned to the correct diagnosis (AVP) was 64.8% for cross-domain consultations.
Conclusion: A novel formulation is now available to deal with problems occurring in a clinical diagnostic decision support system with multi-domain KBs. The utilization of this formulation will help in the development of more integrated KBs with greater focused knowledge domains.

* *Corresponding author at*: 250 Wu-Xin Street, Taipei City 11014, Taiwan. Tel.: +886 2 27361661x2069; fax: +886 2 27387795.
E-mail addresses: drdeung@ms7.hinet.net (Y.-J. Chang), fk002@mail.oit.edu.tw (M.-L. Yeh), amy@ms2.mmh.org.tw (C.-S. Lee), cyh@tmu.edu.tw (C.-Y. Hsu), jack@tmu.edu.tw (Y.-C. Li), wtchiu.tmu@gmail.com (W.-T. Chiu).
URL: http://miew.tmu.edu.tw (Y.-C. Li).
0169-2607/$ – see front matter © 2011 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.cmpb.2011.07.008

1. Introduction

1.1. *Background*

A clinical decision support system (CDSS) is useful in supporting diagnostic and treatment decisions for physicians, especially when facing a myriad clinical symptoms and laboratory test results. A CDSS is also the provision of "clinical knowledge and patient-related information, intelligently filtered or presented at appropriate times, to enhance patient care" [1]. It has been proven that CDSSs improve clinical practice and practitioner performance in 64–68% of the systems [2,3]. Furthermore, computerized physician order entry system with CDSS integration can better improve the quality of health care measured by adherence to recommended guidelines [4]. For medical students, decision support systems may also serve as didactic tools for critical learning [5–7].

CDSS has been categorized as knowledge based and non-knowledge based system by the employment of different theoretical bases. Most of the current knowledge-based CDSS arose from earlier expert systems research where the aim was to build a computer program that could simulate human reasoning with either logic or probability theories. The three parts of most CDSS are the knowledge base (KB), the inference or reasoning engine, and a mechanism to communicate with the user. In the recent twenty years, the developers of these systems have begun to adapt them so that they could be used more easily to support real-life patient care activities [8]. Probability, using Bayes' formulation, is widely applied as it is most similar to medical practice in dealing with uncertainty and with the advantage of assisting in decision making as new information is available.

However, many difficulties are inherent in constructing and maintaining a multi-domain large-scale decision support system. Therefore, it is preferable to build and maintain a system with several smaller knowledge bases, which focus on different knowledge domains. However, these KBs deal with very specific and relative narrow medical specialties. In most clinical situations, domain experts from different specialty background may perceive a given disease differently and thereby giving a different diagnosis. When we build KBs based on these domain experts' respective estimation, we could generate KBs containing the same disease but use different probabilistic estimates. This can result in inconsistent output when we try to integrate answers generated from different KBs.

Yan [9] and his colleagues have proposed an Internet-based knowledge acquisition (KA)/management method to construct large-scale medical KBs. They used an 8-digit numeric coding scheme with weighted value system to implement a clinical decision system that can cover as many as 10^8 diseases. The weighted value for each symptom or manifestation in one disease is classified as four important levels and their respective diagnostic weighted values are suggested to be 0.2, 0.4, 0.6, and 1.0. The determination of the final weighted values is mainly relied on authorized experts. Their work contributes the Internet-based KA method capable of acquiring and managing knowledge in a cost-effective manner, especially for several KBs, but less detailed in the description of the formation of weight value.

1.2. *Cross-domain decision making*

Non-infectious generalized blistering diseases (GBD) and autoimmune diseases (AID) are common in dermatology and rheumatology. Both share common clinical findings and laboratory test results that yield uncertainty in the diagnosis. Dermatologists and rheumatologists often encounter patients in multi-discipline settings. Developing a cross-domain CDSS is important to resolve the issue of cross-domain inference. We used the well-established "Probabilistic Dermatopathological Diagnostic Decision Support System" [10], which contains a KB for GBD already, as the framework in dealing with the uncertainties existing in the diagnosing of disease. For the purpose of this study, researchers built a new KB for AID using knowledge acquisition technology described previously [10]. By means of the mathematical formulation named "cross-domain Bayesian formulation", values of *a priori* evaluation, true positive rate (TPR) and false positive rate (FPR) of the two KBs could be transformed in the different domains.

The purpose of this study is to describe the cross-domain Bayesian formulation that can be used to accommodate the problem stated above. With this new formulation, it is possible to consolidate numerous small and focused KBs that were developed at different locations, and integrate them into a large and multi-domain decision support system with consistency.

2. Methods

There are two main parts in the process of constructing a cross-domain decision support system:

(1) Knowledge representation and system shell.
(2) Cross-domain probabilistic inference.

2.1. *Knowledge representation and system shell*

Knowledge representation is the core of a decision support system, which includes the methodology of knowledge engineering, structure of knowledge, and algorithm of inference engine. Probabilistic inference uses probability to present the uncertainty of a knowledge field; it also uses a mathematical formula to calculate the inference result. The most popular one is the Bayesian formulation. The advantages of multi-membership Bayesian formulation are: (1) it allows "multiple diagnoses" for the same set of findings; (2) it is easier to understand and implement; and (3) it is an easier formulation than the Bayesian network, so that one could compute all relevant diagnosis in relative short time on a personal computer [11–16].

Inference engine, user interface, and other tools used in maintaining the knowledge base make up the "system shell". Our inference engine is based upon a modified Bayesian multi-membership, which will be described in the following test. We used an Internet-based interactive interface due to its friendly and graphical user interface.

In the process of construction of the medical decision support system, knowledge acquisition is the most time- and resource-consuming step. In this process, major human