Knowledge-Based Systems 54 (2013) 128–136

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

A bi-level belief rule based decision support system for diagnosis of lymph node metastasis in gastric cancer

Zhi-Guo Zhou [a,b], Fang Liu [a,b,*], Li-Cheng Jiao [b], Zhi-Jie Zhou [c], Jian-Bo Yang [d], Mao-Guo Gong [b], Xiao-Peng Zhang [e]

[a] School of Computer Science and Technology, Xidian University, Xi'an 710071, PR China
[b] Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education of China, Xidian University, Xi'an 710071, PR China
[c] High-Tech Institute of Xi'an, Xi'an, Shaanxi 710025, PR China
[d] Manchester Business School, The University of Manchester, Manchester M15 6PB, UK
[e] Key laboratory of Carcinogenesis and Translational Research of Ministry of Education of China, Department of Radiology, Peking University School of Oncology, Beijing Cancer Hospital & Institute, Beijing 100142, PR China

ARTICLE INFO

Article history:
Received 27 February 2013
Received in revised form 30 August 2013
Accepted 2 September 2013
Available online 29 September 2013

Keywords:
Clinical decision support system
Gastric cancer
Lymph node metastasis
Belief rule base
Clonal selection algorithm

ABSTRACT

Lymph Node Metastasis (LNM) in gastric cancer is an important prognostic factor regarding long-term survival. As it is difficult for doctors to combine multiple factors for a comprehensive analysis, Clinical Decision Support System (CDSS) is desired to help the analysis. In this paper, a novel Bi-level Belief Rule Based (BBRB) prototype CDSS is proposed. The CDSS consists of a two-layer Belief Rule Base (BRB) system. It can be used to handle uncertainty in both clinical data and specific domain knowledge. Initial BRBs are constructed by domain specific knowledge, which may not be accurate. Traditional methods for optimizing BRB are sensitive to initialization and are limited by their weak local searching abilities. In this paper, a new Clonal Selection Algorithm (CSA) is proposed to train a BRB system. Based on CSA, efficient global search can be achieved by reproducing individuals and selecting their improved maturated progenies after the affinity maturation process. The proposed prototype CDSS is validated using a set of real patient data and performs extremely well. In particular, BBRB is capable of providing more reliable and informative diagnosis than a single-layer BRB system in the case study. Compared with conventional optimization method, the new CSA could improve the diagnostic performance further by trying to avoid immature convergence to local optima.

1. Introduction

Gastric cancer is one of the leading causes for cancer-related death in the world [25]. Lymph Node Metastasis (LNM) is one of the most important prognostic factors regarding long-term survival [7,24,30,34]. The TNM[1] staging system based on American Joint Committee on Cancer (AJCC) is taken as the evaluated standard and has been widely accepted [15]. Based on this standard, the 5-year survival rate of patients in the N0 stage after surgery is 86.1%, while the survival rate for the N1, N2 and N3 stage patients is just 58.1%, 23.3% and 5.9%, respectively [45].

Currently, doctors manually diagnose LNM according to the size of lymph nodes, which relies on various imaging methods, such as multi-slice spiral Computerized Tomography (CT), Magnetic Resonance Imaging (MRI) and Positron Emission computed Tomography (PET). However, any of the above imaging tools cannot assess the lymph node status satisfactorily. For example, large lymph nodes may be caused by inflammation, while small ones may be metastatic. In fact, many studies have shown that LNM is not only associated with the size of lymph nodes but also with the number of lymph nodes. It is difficult for doctors to combine the above two factors to conduct a comprehensive analysis. Fortunately, Clinical Decision Support System (CDSS) from Information Technology (IT) can be used to help solve this problem, which has been applied to medical domains successfully [1,18].

In the current studies, several CDSSs have been proposed for diagnosing LNM, such as Support Vector Machine (SVM) based CDSS [46], and Artificial Neural Network (ANN) based CDSS [9]. However, there are some limitations when these methods are applied in reality. Firstly, these methods are block-box modeling approaches in nature and their internal structures are not directly linked to the reasoning logic or process, which makes it difficult for doctors to know how important each medical attribute is regarding prediction results. Secondly, ANN and SVM are implemented by

* Corresponding author at: School of Computer Science and Technology, Xidian University, Xi'an 710071, PR China. Tel.: +86 029 88204310; fax: +86 029 88201023.
E-mail address: f63liu@163.com (F. Liu).

[1] The international normative TNM classification describes the state of the tumor (T), the lymph nodes (N), and possible metastases (M).

0950-7051/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.knosys.2013.09.001

pure nonlinear data fitting. However, as doctors play an essential role in determining the final stage, it is important to evaluate LNM using both clinical data and doctors' domain specific knowledge.

From the above discussions, it can be concluded that CDSSs can take into account human judgments are more appropriate to help diagnose LNM. For knowledge based CDSSs, representing and reasoning with uncertain medical knowledge are critical areas which need outstanding methodologies and techniques [18,23]. This is because uncertainties exist in clinical signs, clinical domain knowledge and clinical symptoms, and it is difficult to select domain knowledge and construct knowledge bases [11]. Several CDSSs have been developed by using domain knowledge. An early rule-based expert system named as MYCIN [29] was developed in the early 1970s. Another CDSS called Iliad [33] was developed to assist medical decision-making. In Iliad, a sequential Bayesian inference algorithm is used to generate a ranked list of diagnoses and assign a posterior probability to each diagnosis. Fuzzy logic was used in CDSS for renal transplantation assignment [44]. However, the above methods have their deficiencies in knowledge representation and reasoning. Complete knowledge about all parameters is required in a Bayesian inference model. This creates a bottleneck in a knowledge acquisition process that can be extremely time-consuming. As there is usually a fuzzification and de-fuzzification procedure in a fuzzy inference process, reasoning is controversial in most fuzzy logic based systems [11].

To support modeling and inference with clinical domain knowledge under uncertainties, a recently developed belief Rule-based Inference Methodology using the Evidential Reasoning approach (RIMER) [41] is employed for developing an intelligent CDSS [11] which is used for clinical risk assessment of cardiac chest pain. In this CDSS, Belief Rule Base (BRB) is used to model clinical domain knowledge and reasoning is implemented by the Evidential Reasoning (ER) approach [37,38,40–42]. Compared to other CDSSs as described above, it has the following advantages: (1) BRB can be initialized using domain knowledge or assigned randomly, because the belief rules can be adjusted by historical clinical data, which can help circumvent the bottleneck of estimating all initial parameters. (2) Different types of uncertainties can be included in initial BRB as ER can preserve the original features of uncertainties in inference process and the corresponding effects can also be represented in the final conclusion.

According to the clinical domain knowledge, LNM can be divided into four stages. The higher the stage is, the more severe the illness. Stage 0 represents that LNM does not occur, while stage 3 shows that metastasis is most serious. These two stages can be predicted easily by a single-layer BRB system. However, the prediction of stages 1 and 2 is difficult because they are easily confusing [13,17]. So a single-layer BRB system is difficult to diagnose LNM accurately. To solve this problem, a Bi-level Belief Rule Based (BBRB) prototype CDSS is proposed in this paper. The new prototype CDSS consists of two layers [10,16,28] which are modeled by a single-layer BRB. In the first layer, original stages 1 and 2 are integrated into a new stage, resulting in three stages which include original stages 0, stage 3 and the new stage. In the second layer, a specialized BRB is constructed to distinguish stage 1 from stage 2. Compared to single-layer BRB, BBRB can provide more reliable and informative diagnosis for LNM. As manually constructed belief rules may not be accurate, there is a need to train BBRB. However, training a BRB system is sensitive to system initialization and may lead to a locally optimized BRB system. In this paper, a novel Clonal Selection Algorithm (CSA) for training BRB system is proposed. Compared to conventional methods, new CSA is capable of improving performance significantly.

The rest of the paper is organized as follows. The problem formulation is shown in Section 2. In Section 3, the BBRB CDSS prototype is described. A new CSA based method for optimizing BRB is developed in Section 4. In Section 5, the proposed CDSS prototype for diagnosing LNM is presented. The validation of BBRB is discussed in Section 6. This paper is concluded in Section 7.

2. Problem formulation

Suppose that $x = [x_1, x_2, \ldots, x_M]$ is the diagnosis factor which are obtained from medical images such as CT images, where $x_i (i = 1, \ldots, M)$ denotes an attribute and M is the number of attributes. It is assumed that D is the output and P is the corresponding parameter vector. The LNM diagnosis problem is in essence aimed to establish causal relationships between x and D. Such relationships are generally represented by

$$D = f(x, P) \tag{1}$$

where f is in general a function of D. As there are four stages, the output can also be represented as:

$$D = \{(D_0, \beta_0), (D_1, \beta_1), (D_2, \beta_2), (D_3, \beta_3)\} \tag{2}$$

where $D_i (i = 0, \ldots, 3)$ is the stage of LNM, $\beta_i (i = 0, \ldots, 3)$ is the corresponding belief degree and the following constraints are satisfied.

$$\sum_{i=0}^{3} \beta_i = 1 \quad \text{and} \quad 0 \leqslant \beta_i \leqslant 1, \quad i = 0, \ldots, 3 \tag{3}$$

As mentioned in Section 1, D_1 and D_2 is difficult to distinguish. Therefore, a new BBRB model will be presented to solve this problem. On the other hand, the parameter P is initialized by expert knowledge and may be inaccurate. Although the optimal learning methods have been proposed, they are sensitive to initialization and may lead to local optimization. Therefore, a new CSA based method is proposed, which will be described in Section 4.

3. BBRB prototype CDSS

In this section, the BBRB model is presented as a two layer system.

3.1. Proposed BBRB model

The new CDSS prototype is shown in Fig. 1, where $x_1, x_2, \ldots, x_M$ denote the inputs and stage 0–stage 3 is the final output. BRB_1 represents the first layer in BBRB, while BRB_2 represents the second layer. In the first layer, the original stages D_1 and D_2 are integrated into a new stage, and the output O_1 is represented as:

$$O_1 = \left\{\left(D_1^1, \beta_1^1\right), \left(D_2^1, \beta_2^1\right), \left(D_3^1, \beta_3^1\right)\right\} \tag{4}$$

where $D_i^1 (i = 1, \ldots, 3)$ is the output in BRB_1 and $\beta_i^1 (i = 1, \ldots, 3)$ is the belief degree which satisfies:

$$\sum_{i=1}^{3} \beta_i^1 = 1 \quad \text{and} \quad 0 \leqslant \beta_i^1 \leqslant 1, \quad i = 1, \ldots, 3 \tag{5}$$

If the result is D_1^1 or D_3^1, it shows that LNM will be predicted as D_0 or D_3 and the decision process ends. Otherwise, the second layer will be activated. Let O_2 represent the output of BRB_2 as follows:

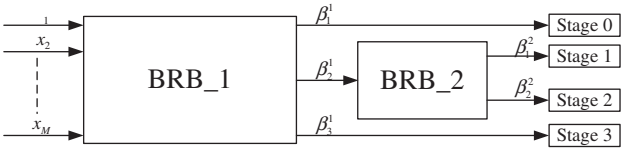

Fig. 1. The new BBRB CDSS prototype.