NEW APPROACH TO TREAT UNCERTAINTY IN DIAGNOSING CARDIOVASCULAR DISEASE BY USING BAYESIAN THEOREM

BOOMA DEVI SEKAR, MINGCHUI DONG, WAI KEI LEI, JUN SHI, JIAYI DOU
Faculty of Science and Technology, University of Macau
Avenida Padre Tomas Pereira, Taipa
MACAU SAR, CHINA
sboomadevi@yahoo.com http://www.boomaprofile.info

Abstract: - A new approach to define and assign statistical parameters to Bayesian inference nodes derived from fuzzy logic technology is proposed. First to develop an intelligent medical diagnostic system, the individual membership function can be pre-defined by matching separately the adapted high-order polynomial, S-type or quasi-Gaussian function with plot of collected clinical diagnostic data. Consequently the coefficients in the defined membership function formulas are fixed which in the diagnostic process can be used to define membership grades vs. recorded symptoms dynamically and individually. Based on symptom-mapped membership grades, statistical parameters can be further defined and assigned to each relevant node in inference nets. The simplicity and adaptability of the proposed methodology is demonstrated and tested by applying it in diagnosing 5 most common and important cardiovascular diseases, through constructed hierarchical Bayesian fuzzy inference nets. The defined statistical parameters are used in calculating propagation of probabilities using Bayesian theorem to solve refractory uncertainty and deduce the disease(s).

Key-Words: statistical parameters; uncertainty; membership function; Bayesian inference nets; propagation of probability, diagnosis of cardiovascular disease

1. Introduction

Ever since the successful implementation of PROSPECTOR [1] Expert system (ES), designed to aid geologists in exploring hidden mineral deposits in the mid 1970s, researchers started investigating the use of Bayesian inference nets [2] in various applications such as document classification [3], bioinformatics [4], information retrieval [5] etc. However, the key difficulty was ‘how to construct such nets?’ and ‘how to define and assign statistical parameters (prior probability, likelihood of belief or certainty factor etc.) to inference nodes?’, which are needed to solve uncertainty and calculate propagation of probabilities. The first issue has already been addressed in our previous paper by proposing a generalized Bayesian inference nets model (GBINM) and applying it to construct inference nets for diagnosing cardiovascular diseases (CVDs), presently under review in a refereed journal. This paper will mainly focus on solving the second issue, defining and assigning statistical parameters.

Literatures show that in well established ES like PROSPECTOR [1] and MYCIN [6], static values were assigned to statistical parameters using subjective (experts’ estimation) method. Since confliictions existed between different expert’s opinions, it causes more complexity to ES designer in defining and assigning the appropriate statistical parameters. Thus, in our proposed approach the dynamic values are defined and assigned to each node of inference nets through mapping symptoms to membership grades on individual well defined membership function, defined by using both on-site measured records and experts’ estimation. Such a novel method has been successfully applied and tested in motor fault diagnostic Bayesian inference nets [7]. Expanding from non-life to alive problems, in this paper, its availability and robustness are further checked in much more complicated application, diagnosis of CVDs. Five important and most common types of CVDs, namely coronary heart disease (CHD), hypertension (HT), arrhythmia (AR), pulmonary heart disease (PHD) and cerebral infarction (CIN) are diagnosed by using physiological messages and hemodynamic parameters (HDPs), obtained through a non-invasive sphygogram (SPG) analysis.

2. Define and Assign Statistical Parameters

In any designed ES, statistical parameters play key
A role in addressing uncertainty. Well known expert system like PROSPECTOR uses prior probability and likelihood of belief (likelihood of sufficiency (LS), likelihood of necessity (LN)) and MYCIN.
uses certainty factor (CF) as statistical parameters to deal with uncertainty. However, the major difficulty that persecutes researchers often is to define and assign these statistical parameters correctly. As discussed above, the normally practiced subjective approach of assigning static values to statistical parameters introduced more uncertainty. Thus the paper with a specific example, using hierarchical Bayesian fuzzy inference nets (HBFIN) to diagnose CVDs, will introduce a novel approach derived from FL technology to define and assign dynamic values of statistical parameters.

2.1. Hierarchical Bayesian Fuzzy Inference Nets with Assigned Statistical Parameters

Fig. 1 shows the partially constructed HBFIN with statistical parameters defined and assigned to each inference node while input a medical record sampled from a patient, eventually confirmed of having hypertension (HT). Here the structure of HBFIN is constructed by using GBINM [7]. The needed medical knowledge is based on the analysis of 1190 clinical records and 700 healthy records collected from 165 patients within 5 to 8 weeks in two hospitals of China. A sampled medical record includes each patient’s 6 physiological messages (age, sex, height, weight, systolic and diastolic blood pressure), original SPG data, 32 HDPs [8] and doctor’s clinical diagnostic results. The step by step process of analyzing the medical records and constructing HBFIN using GBINM is elaborately explained in our previous paper mentioned earlier.

In HBFIN, the prior probability is assigned on up-left corner outside of each inference node and LS, LN are assigned beside the arc connecting two inference nodes. To reduce the complexity of nets the statistical parameters are defined and assigned only to those significant inference nodes. For inference nodes executing conjunction and disjunction operation, the statistical parameters would be later calculated from the evidences contributing to these nodes.

The following section will discuss how these statistical parameters are defined and assigned automatically in diagnostic process using the pre-defined fuzzy logic (FL) membership functions for all m diseases vs. n symptoms.

2.2 Define Statistical Parameters using Fuzzy Logic Membership Function

Various types of function such as Gaussian, triangle, high-order polynomial, S-type etc., can be used to define the membership functions depending on data distribution. Based on the analysis of doctors’ estimate and histogram obtained from the frequency plot of observed medical records, high-order polynomial, S-type or quasi-Gaussian functions are adapted to define the individual membership function.

The general formula of i-th membership function $f_i(s_j^{(0)})$ vs. j-th symptom in 0-th stage $s_j^{(0)}$ expressed in high-order polynomial, S-type and quasi-Gaussian function is listed sequentially below:

$$f_i(s_j^{(0)}) = \lambda_0 + \lambda_1 s_j^{(0)} + \lambda_2 (s_j^{(0)})^2 + ... + \lambda_t (s_j^{(0)})^t$$ (1)

where $t$ is the order of polynomial, $\lambda_0$ is random error or noise component, and $\lambda_1$, $\lambda_2$, …, $\lambda_t$ are coefficients.

$$f_i(s_j^{(0)}) = \frac{1}{1+e^{-a(s_j^{(0)}-b)}}$$ (2)

where $b$ is the turning point of curve and $a$ is the slope of function.

$$f_i(s_j^{(0)}) = e^{-\left(\frac{s_j^{(0)}-a_i(s_j^{(0)})}{2d_i(s_j^{(0)})}\right)^2}$$ (3)

where $a_i(s_j^{(0)})$ is the maximum membership grade and $2d_i(s_j^{(0)})$ is the bandwidth of that function.

Here the word ‘quasi’ is expressed to indicate that the membership function plot would appear as Gaussian distribution shape, but will not cover to the extent of positive and negative infinity.

As an example, Fig. 2 demonstrates the generation of high-order polynomial or quasi-Gaussian membership function for the input symptom ‘Age’ based on the analysis of doctors’ estimate (expressing the likelihood of having certain CVD at different age) and histogram with contour curve obtained from frequency plot of observed 1190 records of patients suffering from certain CVD.

![Fig. 2 Generation of high-order polynomial or quasi-Gaussian membership function for certain CVD vs. input symptom ‘Age’](image)
Similarly the membership functions for all \( m \) CVDs vs. \( n \) symptoms are constructed and shown in Fig. 3.

Bigger \( \alpha \) would increase probability of the hypothesis to be true in presence of the evidence, meanwhile smaller \( \beta \) would increase probability of the hypothesis to be false in absence of the evidence. In this paper, \( \alpha = 100 \) and \( \beta = 10 \) are experimentally chosen. The statistical parameters for the intermediate hypothesis or conclusion nodes are defined and assigned using the principle of indifference [9].

Now as the statistical parameters are defined and assigned automatically using the proposed methodology, next to diagnose CVDs using HBFIN, these statistical parameters and Bayesian theorem will be further used to calculate the propagation of probabilities.

### 2.3. Propagation of Probability

The Bayesian inference nets form a static knowledge structure, in which the probability associated with each functional/hypothesis node consequently changes when the evidence is certain or uncertain. This change in probability is propagated up stage by stage through the hierarchical Bayesian inference nets to ultimately support or disprove the top-level hypothesis/conclusion. Eqs. (4) ~ (8) [2] are used to calculate the propagation of probability.

Here, when the evidence contributing to a particular hypothesis node is known as true or false, it represents certain evidence. When the evidence is unknown, partially known, or not quite sure etc., it represents uncertain evidence. However, in this paper conditional independence of the evidence is assumed. Therefore for partially known or not quite sure evidence, according to its degree of belief, it is categorized to be true or false and the propagation of probability is calculated accordingly.

Prior odds of \( k \)-th hypothesis on \( q \)-th stage:

\[
O(h^w_k) = \frac{P(h^w_k)}{1-P(h^w_k)}
\]

(4)

where \( P(h^w_k) \) = Prior probability of \( k \)-th hypothesis on \( q \)-th stage.

Posterior odds of \( k \)-th hypothesis on \( q \)-th stage:

\[
O(h^w_k | x \text{ (or) } e) = \prod_{i=1}^{n} L_i O(h^w_k).
\]

(5)

here

\( x = \) Certain evidence to the \( k \)-th hypothesis on \( q \)-th stage that is known as true;

\( e = \) Uncertain evidence to the \( k \)-th hypothesis on \( q \)-th stage;
L_4 = LS_i \{ \text{for all evidences to the k-th hypothesis on q-th stage that are known as true}; \\
L_4 = LN_i \{ \text{for all evidences to the k-th hypothesis on q-th stage that are known as false}; \\
L_4 = 1 \{ \text{for all evidences to the k-th hypothesis on q-th stage that are unknown}. \\

Posterior probability of k-th hypothesis on q-th stage when the evidence is certain:

\[
P(h_{i}^{q}|x) = \frac{O(h_{i}^{q}|x)}{1+O(h_{i}^{q}|x)} (6)
\]

In a multilayer Bayesian inference nets, the intermediate hypothesis node on (q-t)-th, q = 1, 2, ..., n; t = 0, 1, 2, ..., q-1, may also act as the evidence to generate the hypothesis node on the q-th stage. The prior probability for this node on q-th stage becomes inconsistent/uncertain with the probability derived from its antecedents. Different methods have been proposed to solve this inconsistency/uncertainty. In this paper when the evidence is uncertain, Eqs. (7) – (8) [2], [10] derived using piecewise linear interpolation function (an ad-hoc method) shown in Fig. 4 [2], [10] are used to calculate the posterior probability of hypothesis/conclusion node.

\[x' = \text{Certain evidence to the k-th hypothesis on q-th stage that is known as false};
\]
\[P(x') = \min[P(x_i^{(0)}),...P(x_i^{(q)}),P(h_i^{(q+1)}),...P(h_i^{(q+q-1)}), P(h_i^{(q)}), P(F_i^{(q+q+1)}),...]
\]
for conjunction operation;
\[P(x') = \max[P(x_i^{(0)}),...P(x_i^{(q)}),P(h_i^{(q+1)}),...P(h_i^{(q+q-1)}), P(h_i^{(q)}), P(F_i^{(q+q+1)}),...]
\]
for disjunction operation.

3. Testing Results
The function and validity of constructed HBFIN are examined using the site-measured extra patients’ 367 medical records and 150 healthy records, which are partially shown in Table 1. The diagnostic results of HBFIN are shown in Table 2.

Table 1. Partial Medical Records Sampled from Five Patients

<table>
<thead>
<tr>
<th>* Input Symptoms (units)</th>
<th>Patients’ partial medical records</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (yrs)</td>
<td>56  50  38  60  74</td>
</tr>
<tr>
<td>BMI (kg/m’2)</td>
<td>28.9 29.2 28 30.1 25.6</td>
</tr>
<tr>
<td>SBP (mm Hg)</td>
<td>160 120 170 130 164</td>
</tr>
<tr>
<td>DBP (mmHg)</td>
<td>90  90  70  75  96</td>
</tr>
<tr>
<td>CI (ml/stroke/ min)</td>
<td>4.50 2.63 5.37 5.40 4.88</td>
</tr>
<tr>
<td>FEK (no unit)</td>
<td>0.35 0.24 0.17 0.51 0.19</td>
</tr>
<tr>
<td>PP (mm Hg)</td>
<td>70 30 100 55 68</td>
</tr>
<tr>
<td>MDP (mmHg)</td>
<td>108.6 94.3 89.3 88.9 111.9</td>
</tr>
<tr>
<td>CMBV (ml/ min)</td>
<td>460.7 345.2 236.9 675.4 376.1</td>
</tr>
<tr>
<td>η (cp)</td>
<td>2.47 5.67 1.76 2.60 2.71</td>
</tr>
<tr>
<td>MAP (mm Hg)</td>
<td>129.6 101.4 115.6 105 131.3</td>
</tr>
<tr>
<td>MSP (mm Hg)</td>
<td>150.6 108.6 142.0 121.2 150.6</td>
</tr>
<tr>
<td>HOV (ml/ min)</td>
<td>50.29 42.29 63.39 34.56 58.40</td>
</tr>
<tr>
<td>PR (beats/ min)</td>
<td>59.85 74.94 64.81 59.13 84.50</td>
</tr>
<tr>
<td>CCP (mm Hg)</td>
<td>65.87 81.09 30.98 66.94 80.12</td>
</tr>
<tr>
<td>PAWP (mmHg)</td>
<td>24.12 8.90 39.01 8.05 15.87</td>
</tr>
</tbody>
</table>

Table 2. Diagnostic Results of HBFIN

<table>
<thead>
<tr>
<th>Patients</th>
<th>Probability of having certain CVD</th>
<th>Doctor’s diagnosed result</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HT</td>
<td>CHD</td>
</tr>
<tr>
<td>1</td>
<td>0.95</td>
<td>0.63</td>
</tr>
<tr>
<td>2</td>
<td>0.30</td>
<td>0.86</td>
</tr>
<tr>
<td>3</td>
<td>0.32</td>
<td>0.02</td>
</tr>
<tr>
<td>4</td>
<td>0.94</td>
<td>0.64</td>
</tr>
<tr>
<td>5</td>
<td>0.91</td>
<td>0.48</td>
</tr>
</tbody>
</table>

Although Table 2 shows the testing result of 5 patients only, equally good results are obtained for rest of the testing medical records. Considering that in the proposed non-invasive approach, the diagnosis depends only on the strongly physically
related parameters HDPs, such diagnostic accuracy is highly acceptable, which proves the suitability of the hybrid intelligent diagnostic system for home healthcare usage. Our other testing results currently under research show that after adding the blood-testing symptoms offered by invasive approach, the diagnosis accuracy of HBIFN could be further improved.

The sensitivity of diagnosis result to coefficients (α and β used to define LS and LN in Rules 1~2) tuning has also been tested and the results are presented in Table 3. The medical record sampled from patient 1 in Table 1 is used for this study.

Table 3. Testing Results of Coefficient Tuning

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Probability result of patient with confirmed HT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HT</td>
</tr>
<tr>
<td>10 10</td>
<td>0.62</td>
</tr>
<tr>
<td>20 10</td>
<td>0.79</td>
</tr>
<tr>
<td>30 10</td>
<td>0.85</td>
</tr>
<tr>
<td>50 10</td>
<td>0.91</td>
</tr>
<tr>
<td>100 1</td>
<td>0.95</td>
</tr>
<tr>
<td>100 5</td>
<td>0.95</td>
</tr>
<tr>
<td>100 20</td>
<td>0.95</td>
</tr>
<tr>
<td>100 30</td>
<td>0.95</td>
</tr>
<tr>
<td>100 50</td>
<td>0.95</td>
</tr>
</tbody>
</table>

It can be observed that the bigger α increases the probability of hypothesis to be true in presence of the evidence, meanwhile the smaller β increases the probability of hypothesis to be false in absence of the evidence.

4. Conclusion
An attempt to solve the difficulty of defining and assigning statistical parameters to Bayesian inference nodes which persecutes the researchers often is presented. The values of statistical parameters are defined and assigned to relevant inference nodes by mapping input symptoms to the pre-defined membership functions dynamically and individually. The simplicity and adaptability of proposed approach is illustrated by applying it in HBIFN to diagnose 5 important types of CVDs. The testing results show high diagnostic accuracy and prove the validity of the proposed methodology.

Acknowledgement:
We would like to thankfully acknowledge the financial support of the Macau Science and Technology Development Foundation for the project under Grant Number: 018/2009/A1.

References: