Oriental Medical Data Mining and Diagnosis Based On Binary Independent Factor Model

Jeong-Yon Shim †, Lei Xu ‡

†Dept. of Computer Software, YongIn SongDam College

MaPyeong Dong 571-1, YongIn Si, KyeongKi Do, Korea

‡Department of Computer Science and Engineering

Chinese University of Hong Kong,Shatin, NT,Hong Kong, China

Phone 852 2609 8423, Fax 852 2603 5024

Abstract

In an oriental medical diagnostic process, human doctor accumulates the observed complex symptom data from the patient and diagnoses the person's illness. The doctor tries to infer from the visible observed symptoms based on his knowledge and what his understanding before he comes to a conclusion. A complicated observation is regarded as generated from a number of hidden and simpler factors and can be analyzed by a factor analysis model. We develop an automatic oriental medical diagnostic system which consists of a specific three layer model in help of a binary independent factor analysis based on BYY harmony learning. We consider the diagnosis for a kind of the oriental causal illness name, 'GanDamSpYeol', with the observed symptoms by this model.

1 Introduction

An important domain for data mining is extracting the diagnostic rules from the patient's data for the medical diagnosis. Human doctor accumulates the observed complex symptom data from the patient and diagnoses the person's illness. The doctor tries to infer from the visible observed symptoms based on the rules in his knowledge base and what his understanding before he comes to a conclusion.

Efforts of building an automatic medical diagnostic system have been made in past years on developing knowledge base expert systems with IF-THEN type production rules for knowledge representation. The study for the medical diagnostic system can be traced back to MYCIN as a well known system. Using production rules and Certainty factor it is successfully developed. However, efforts along this direction suffer certain intrinsic limitations. First, the rules as well as the corresponding reasoning structure are built via manually encoding the knowledge of experienced doctors. It not only takes high cost to obtain the rules from the experts but also has a difficulty in establishing a good rule base. It has a lack of extracting the rules automatically. Second, the system doesn't have automatic learning process and is not easy to be updated adaptively. Because of static property, it has the difficulty that the rule base of this system has to be updated by the experts manually whenever its environment is changed. Moreover, its rule base is increased to the large amount and becomes complex because it has to contain all the possible rules. It may have the problem of redundancy.

As the interest on the oriental medical diagnosis is increased, efforts on building the oriental medical system have been made via the production systems such as in MYCIN. However, in the oriental medical diagnostic domain, the new knowledge processing paradigm and the efficient data mining technology is more necessary because the oriental medical diagnostic process is highly heuristic and the performance of a medical doctor highly depends on his experience. In a general point of view, oriental medical doctor accumulates many samples from the patients in the particular domain, separates several simpler independent factors, classifies them into the several groups and forms rules in his brain. The complex observed symptoms are regarded as generated from the number of simpler factors in a noisy environment.

It is well known philosophy that a complicated observation is regarded as generated from a number of hidden and simpler factors. Many efforts in this area have also been made on developing mathematical theories that implement this philosophy in a number of scientific and engineering fields. The earliest effort can be traced back to the beginning of the 20th century by Spearman [1], and had been followed by various studies in the literature of statistics and neural network, which uses a linear model $x = Ay + e$ that interprets the observed data $x = [x^{(1)}, ..., x^{(d)}]^T$ as generated from hidden factors $y = [y^{(1)}, ..., y^{(k)}]^T$ via a linear system disturbed by certain noise e that is independent of y.

One recent approach for implementing such a concept is an *independent factor model* based on Bayesian Ying Yang (BYY) harmony learning, proposed by one of authors. BYY Independent Factor Model can be applied to the oriental medical diagnostic area by providing a new tool for separating the independent factors from the observed mixed data. Its ability of adaptive learning and of determining number of independent factors makes it a more appropriate method than the production rule based model in the doctor's heuristic experience based oriental medical domain.

In this paper, we develop an oriental medical diagnostic system based on the BYY binary independent model implemented in a three layer structure. In Sec.2, BYY independent factor model is introduced. In Sec.3, an automatic oriental medical diagnostic system is presented in a three layer diagnostic model in help of an independent binary factor model based on BYY harmony learning. In Sec. 4, we apply it to the diagnostic problem for the oriental causal illness from the observed symptom data.

2 Independent Binary Factor Model via BYY harmony learning

2.1 Independent Binary Factor Analysis

We consider the linear model $x = Ay + e$ with a Gaussian noise e of zero mean and covariance matrix Σ as well as $Eye^T = 0$. Equivalently, such a linear model can be described by the following distribution

$$q(x|y, \phi) = G(x|Ay, \Sigma),\ \phi = \{A, \Sigma\}. \quad (1)$$

where $G(u|m,\ \Sigma)$ means a Gaussian distribution of variable u with mean m and covariance matrix Σ.

Moreover, we consider $y = [y^{(1)}, ..., y^{(k)}]^T$ with each component subject to a Bernoulli distribution. That is, we have

$$q(y|\psi) = \prod_{j=1}^{k} q_j^{y^{(j)}} (1 - q_j)^{1-y^{(j)}}, \psi = \{q_j\}_{j=1}^{k}, \quad (2)$$

where the dimension k indicates the scale of the inner representation y. In this case, k is simply the number of binary factors. If k are too small , we could not get good inner representations due to the limited representing capacity. On the other hand , if they are too large, it is difficult to form compact representations thus the modeling performance is poor.

Both $q(x|y, \phi)$ and $q(y|\psi)$ together specify an independent binary factor model. Assume that we have a set of samples of x, our learning task consists of specifying the parameter set $\theta = \{\phi, \psi\}$, which is called parameter learning, and specifying the number k of factors, which is called model selection in the sense that different values of k correspond different models with a same configuration but in different scales.

2.2 BYY harmony learning for Independent Factor Analysis

The Bayesian Ying-Yang (BYY) learning was proposed as a unified statistical learning framework by one of the present authors firstly in 1995[6] and systematically developed in past years. The obtained results can be summarized from both the perspective of general learning framework and the perspective of specific learning paradigms. From the 1st perspective, BYY learning consists of a general BYY system and a fundamental harmony learning principle as a unified guide for developing new regularization techniques, a new class of criteria for model selection, and a new family of algorithms that perform parameter learning with automated model selection. From the 2nd perspective, the BYY learning on specific structures lead us to three major statistical learning paradigms, with a number of new results on the existing major tasks of unsupervised learning and supervised learning, as well as their temporal extensions for modeling data that has temporal relationship among samples[4]. The details results from both the perspectives have been recently summarized in[2,3].

The key idea of Bayesian Ying Yang system is to consider the joint distribution of x, y via two types of Bayesian decomposition of the joint density $q(x|y)q(y) = q(x, y) = p(x, y) = p(y|x)p(x)$. Without any constraints, the two decompositions should be theoretically identical. However, the four components in the two decompositions are usually subject to certain structural constraints. For example, in our case we have $q(y) = q(y|\psi)$ with the structure in eq.(2) and $q(x|y) = q(x|y, \theta)$ with the structure in eq.(1). Moreover, we can also let $p(x)$ given by a nonparametric estimate from samples of x. Similar, $p(y|x)$ may also has a specific structure.

Thus, we usually have two different but complementary Bayesian representations:

$$\begin{aligned} p(x, y) &= p(y|x)p(x), \\ q(x, y) &= q(x|y)q(y), \end{aligned} \quad (3)$$

which, as discussed in the original paper[6], compliments to the famous Chinese ancient Ying-Yang philosophy with $p(x, y)$ called Yang model that represents the observation domain or space X (or called Yang space) with a underlying $p(x)$ and the forward pathway (or called Yang pathway) by $p(y|x)$, and with $q(x, y)$ called Ying model that represents the invisible inner domain or space Y (or Ying space) by $q(y)$ and the Ying (or backward) pathway by $q(x|y)$. Thus, such a pair of Ying-Yang models is called *Bayesian Ying-Yang (BYY) system.*

Usually, $p(x)$ is given by a nonparametric estimate from samples of x. Here, we simply consider

$$p_0(x) = \frac{1}{N} \sum_{t=1}^{N} \delta(x - x_t). \quad (4)$$

A combination of the structures for the rest three components $q(y)$, $q(x|y)$, and $p(y|x)$ is referred as a system architecture. There are three typical structures. Among them, a backward architecture consists of a parametric $q(x|y)$ while $p(y|x)$ remains free to be decided via learning. In this paper, $q(x|y) = q(x|y, \theta)$ is a parametric model in eq.(1) and we also let $p(y|x)$ free to be decided via learning. That is, we consider a specific backward architecture.

Given two p, q in known structures with each of them having some unknowns in either or both of the scale and the parameters, the task of learning is to specify all the unknowns from the known parts of both the densities. Our *fundamental learning principle* is to make p, q be best harmony in a twofold sense:

- The difference between p, q should be minimized.
- p, q should be of the least complexity.

Mathematically, we use a functional $H(p\|q)$ to measure the degree of harmony between p and q, which is called harmony measure. In our case, the harmony measure takes the following simplified form

$$H(p\|q) = \frac{1}{N}\sum_{t=1}^{N}\sum_{y}p(y|x)\ln\left[q(x|y,\phi)q(y|\psi)\right] - \ln z_q \quad (5)$$

where z_q is a term that has three typical representations. The simplest case is $z_q = 1$ and correspondingly the learning is called empirical learning. This paper considers this simple case only. The other two cases are referred to [2].

In a summary, we can describe a mathematical implementation of the BYY harmony learning as follows

$$\max_{\theta,k} H(p\|q), \quad (6)$$

which can be implemented either in a parallel way such that model selection is made automatically during parameter learning or in a sequential way such that model selection is made after parameter learning. The details are referred to [2,3].

Setting $z_q = 1$ and maximizing $H(p\|q)$ with respect to a free $p(y|x)$, we get

$$\begin{aligned}
&p(y|x) = \delta(y - \hat{y}),\\
&\hat{y} = arg\max_{y}[q(x|y,\phi)q(y|\psi)],\\
&H(p\|q) = \frac{1}{N}\sum_{t=1}^{N}\ln\left[q(x|\hat{y}_t,\phi)q(\hat{y}_t|\psi)\right]. \quad (7)
\end{aligned}$$

Furthermore, maximizing $H(p\|q)$ with respect to ϕ, ψ, we get the following adaptive updating rules

$$\begin{aligned}
&q_j = \frac{1}{1+e^{-\xi_j}},\\
&\xi_j^{new} = \xi_j^{old} + \eta(y^{(j)} - q_j),\\
&e = x - A^{old}\hat{y},\ A^{new} = A^{old} + \eta e\hat{y}^T,
\end{aligned}$$

$$\Sigma^{new} = (1-\eta)\Sigma^{old} + \eta ee^T, \quad (8)$$

where $\eta >$ is a learning step size. Thus, an adaptive learning can be made by repeating getting $\hat{y}$ by eq.(7) and then making updating by eq.(11). The above algorithm eq.(7) and eq.(11) is firstly given in [5] and then further refined with new features in [2].

2.3 Parameter learning with automated model selection

Model selection, i.e., to deciding the number of k hidden independent factors, is an important issue in building an appropriate independent binary factor model. As shown in [2,5], one advantage of the BYY harmony learning is that k can be determined automatically during parameter learning. To realize this point, what needs to do is set k large enough and implement parameter learning by

$$\max_{\theta} H(\theta), H(\theta) = H(\theta, k), \quad (9)$$

during which θ tends to a value θ^* such that the resulted BYY system is equivalent to having an appropriate scale $k^* < k$.

In our problem, the above eq.(9) is implemented by repeating getting $\hat{y}$ by eq.(7) and then making updating by eq.(11). When k is initialized at a large value, during learning, one or more of q_j will tend to either 1 or 0, which means that the corresponding factor y_j can be discarded. So, by setting a threshold of a small positive number ϵ, we simply delete a factor y_j when its corresponding $p_j < \epsilon$ or $p_j > 1 - \epsilon$ during learning.

Alternatively, we can also make parameter learning and model selection sequentially in two steps. With k prefixed, we enumerate k from a small value incrementally, and at each specific k we perform parameter learning by repeating getting $\hat{y}$ by eq.(7) and then making updating by eq.(11), resulting in the best parameter value θ^*. Then, we select a best k^* by

$$\min_{k} J(k), J(k) = -H(\theta^*, k), \quad (10)$$

If there are more than one values of k such that $J(k)$ gets the same minimum, we take the smallest.

In our problem, we have

$$\max_{k} J(k) = -0.5d\log\sigma^2 + \Sigma_{j=1}^{m}[q_j\log q_j + (1-q_j)\log(1-q_j)] \quad (11)$$

Particularly, if we fix $q_j = 0.5$, we have

$$J(k) = -0.5d\log\sigma^2 + k\log 2 \quad (12)$$

3 An oriental medical diagnostic data mining by BYY Binary Independent Factor Analysis

We consider to develop an automatic oriental medical diagnostic system in help of a binary independent factor model based on the BYY harmony learning. Specifically, we regard that the diagnostic process by oriental medical experts actually consists of two steps. First, a long period accumulation of previous experiences on observing a large number of symptoms has developed a factor model x=Ay+e in the brain of experts and thus upon observing the symptoms of a particular patient, the brain removes the noise in the observed symptoms and then associate them to the hidden factors. Second, a set of rule $y \to z$ is set up from these hidden factors to the corresponding diagnostic decision.

This automatic oriental medical diagnostic system has the three layer structure performed by two step algorithm. In the first step algorithm , the binary factor analysis using BYY harmony learning is performed and in the second step algorithm, a set of rules are set up according to their associative relations of $y \to z$ by the learning algorithm.

At Step I, what we need is to learn the factor model $x = Ay + e$ from a training set of symptoms of past patients. We consider a slightly simplified version of eq.(1) with $\Sigma = \sigma^2 I$, i.e., $q(x|y, \phi) = G(x|Ay, \sigma^2 I)$. Still, we let the binary factor $y = [y^{(1)}, ..., y^{(k)}]^T$ given by eq.(2).

From the binary factor analysis, we can extract the rules and interpret by analyzing two terms. One is to analyze the structure and the other is to interpret the mixing matrix, A.

First, we can analyze the structure of y after the best k^* is selected by model selection during the parameter learning,. Specifically, each $y^{(j)}$ is binary taking either '1' denoting that this factor affects an observed symptom or '0' for denoting that this factor is irrelevant to the symptom. Adopting this binary independent factor model in our diagnostic system is justified because it is quite nature to believe that hidden factors should be of the least redundancy among each other, otherwise a compound factor can be further decomposed into simpler factors. Also, it is reasonable to believe an observed symptom resulted from whether one or particular hidden factors occurs, with '1' denoting the occurrence of such a factor and '0' denoting not. Each $y^{(j)}$ is distributed by their associative relations inside pattern.

Assuming the observed groups, $\{P_1, P_2, ..., P_m\}$, the binary factor
$y = [y^{(1)}, ..., y^{(k)}]^T$ for the same group, P_i, selected by the learning algorithm has the similar structure. They are clustered to the similar pattern of binary factors which belong to the same group. By analyzing the structure of y, the medical rules among the patterns can be extracted. We regard the specific structure of y as one rule for its corresponding group. Moreover, in the help of the expert the meanings of each binary factor $y^{(i)}$ can be also extracted and interpreted. Then, this system can provide the appropriate diagnostic tools for the oriental medical diagnosis by extracting not only the medical rules but also hidden causes via binary factor analysis based on the BYY harmony learning and analyzing the structure of y. This technology is very useful in the oriental medical domain where the oriental medical doctor regards the disease as the broken state of harmony in the human body and he tries to find the hidden causes inducing the broken state because it can provide a good tool for developing the oriental medical diagnostic system.

Second, the mixing matrix A is the generative rule generating the symptoms from the hidden factors, y. The generative rules can be induced by interpreted binary factors y and mixing matrix A.

Step I can be adaptively implemented by iterating eq.(7) and eq.(11), which are summarized as follows:

As each x_t comes, we implement three steps iteratively until it converges.

STEP I

(1) Get $y_t = arg\,max_y[q(x_t|y)q(y)]$ or
$y_t = argmin_y\{0.5(\frac{\|x-Ay\|}{\sigma^2})^2 - \sum_{j=1}^{k} y^{(j)} lnq_j + (1 - y^{(j)}) ln(1 - q_j)\}$.

(2) Get $e_t = x_t - A^{old} y_t$,
update $A^{new} = A^{old} + \eta e_t y_t^T$,
$\sigma^{2\ new} = (1 - \eta)\sigma^{2\ old} + \eta \|e_t\|^2$.

(3) Let $q_j = 1/[1 + exp(-c_j)]$,
update $c_j^{new} = c_j^{old} + \eta(y_t^{(j)} - q_j)$;

where $\eta > 0$ is a small step size.

At Step II, we set up another mapping by

$$p(z|y, B, b) = \prod_{j=1}^{m} p_j^{z^{(j)}} (1 - p_j)^{1-z^{(j)}},$$
$$p_j = 1/(1 + e^{-u_j}),$$
$$u = [u_1, \cdots, u_m]^T, \ By + b = u. \qquad (13)$$

It can be made by getting a set of training pair y, z with y obtained from Step I by the mapping $x \to z$ and with z obtained from the corresponding diagnostic records of diagnostic decision by an oriental medical expert.

The B, b can be learned via the maximum likelihood learning adaptively, which is summarized below:
We implement two steps iteratively with the gained $\hat{y}$ from Step I and the corresponding teaching signal z, until it converges:

STEP II

(1) Get $u = [u_1, \cdots, u_m]^T, \ B\hat{y} + b = u$,
$p_j = 1/(1 + e^{-u_j})$,
$f = [f_1, ., f_m]^T, \ f_j = z_j - p_j$

(2) update $B^{new} = B^{old} + \eta f \hat{y}^T$,
$b^{new} = b^{old} + \eta f$,

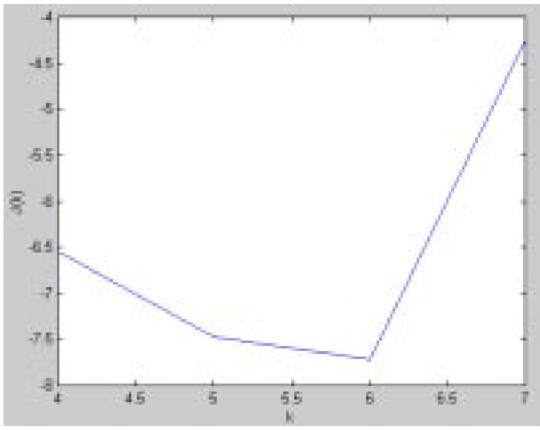

Figure 1: The curve J(k) for source number detection, correct source number $k^* = 6$

4 Experiments

We applied the above three layer model to the diagnostic problem for finding the hidden causal factors from the patient's symptoms and for classifying the current disease. The symptom data of liver disease,oriental causal illness name 'GanDamSpYeol' and 'GanGiBuJog', are used for experiments. We used the accumulated 200 samples of patient including the normal cases for experiment. The data of patients consist of the number $d = 7$ of the observed symptoms and the number $m = 2$ of the diagnostic illness name. The symptoms used in the experiments are denoting 'jaundice', 'fever', 'nausea','yellow urine','taut and rapid pulse','diziness' and 'sallow complexion'. We used two types of disease names for classifying. One is the diagnostic causal disease 'GanDamSpYeol',which is the causal disease denoting the illness state caused by the high dampness and the high heat in the liver and gallbladder in the oriental medical diagnostic area. The other is 'GanGiBuJog' denoting the illness state caused by the deficiency of liver-blood and the deficiency of liver-Gi.

In order to implement this diagnostic process, we use the two step adaptive algorithm given the previous section. In the experiment, the number of $d = 7$ of the observed symptoms are mapped to the selected number of $k = 6$ of factors by model selection. The results of the detection of k^* by $J(k)$ given by equation (11) are shown in figure 1.The minimum point k=6 is selected as the correct source number.

Figure 2 shows the hamming distance errors as the learning of STEP I goes with epoch t and Figure 3 shows the errors of the learning STEP II. After $t = 17$ epochs, the error of Step I converges and after about $t = 70$ epochs, the error of Step II converges. In Step I, the patient's symptom data x_t are recovered to y_t. In Step II the relations between the recovered terms y_t and diagnostic illness name z_t are set up.

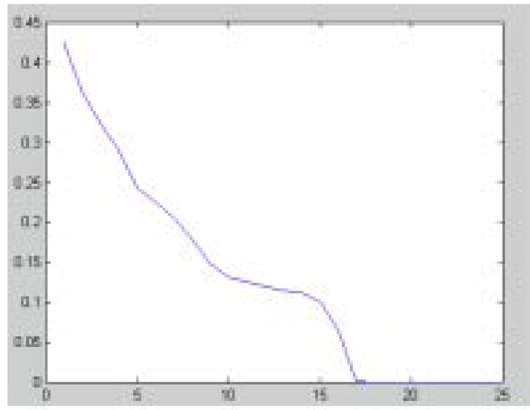

Figure 2: STEP I : Error Curve

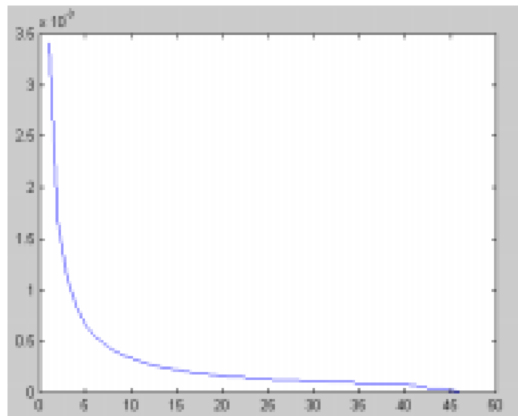

Figure 3: STEP II : Error Curve

In this system, the diagnostic process is performed after the training phase is finished. we get the diagnostic result z_t from the observed symptoms x_t by Step I and Step II sequentially. We select three samples as examples of diagnosis of this oriental medical system for plotting and use the linguistic terms of symptom denoting 'more serious', 'serious', 'medium', 'slight', 'less slight' and 'not' for the purpose of interpretation.

Figure 5 denotes the example of 'GanDamSpYeol', the sample no. 42. The output values of Z_m are produced from 7 symptoms. The values of P_j are 0.9985 and 0.001183. We can make a decision from this values that the diagnostic result is 'GanDamSpYeol' because they are interpreted to $(Z_1 = 1)$ and $(Z_2 = 0)$. Figure 6 shows the example of 'GanGiBuJog', the sample no. 14. From the diagnostic result, we can find the diagnostic disease of this sample is 'GanGiBuJog' because P_j, 0.050552 and 0.987176, is interpreted to $(Z_1 = 0)$ and $(Z_2 = 1)$.

As a result, from testing with 200 sample data of the patients and comparing its diagnostic output data to the diagnostic results of a human doctor, we could get the accuracy of 91.5% though the oriental

```
N = 200
d = 7
no. of source number : 6

estimated A
    0.996794    0.501428    0.439872    0.563472    0.369195    0.299581
    0.406295    0.979532    0.413873    0.464192    0.326245    0.456482
    0.476456    0.510028    0.955932    0.557521    0.572108    0.558098
    0.640172    0.325821    0.331511    1.006564    0.384355    0.286168
    0.387044    0.413482    0.323680    0.409865    0.988121    0.562659
    0.563083    0.384521    0.348309    0.392484    0.571451    1.012098
    0.423355    0.413356    0.282626    0.466111    0.402702    0.465739

estimated B
   -0.778609   0.937206   0.202838   -3.342382   -3.245377   -0.568183
   -0.729695   0.562125   0.952316    3.558773    2.661888   -0.503568
```

Figure 4: Estimated A,B

```
------------------------------------------------------------
The Medical Diagnostic result for the observed symptoms
------------------------------------------------------------

Sample 42

No.             Symptom              Linguistic Degree      value

 1                      jaundice      more serious       3.115630
 2                         fever      more serious       2.991870
 3                        nausea      more serious       3.632000
 4                  yellow urine      more serious       2.867010
 5          taut and rapid pulse           serious       1.800590
 6                     dizziness            slight       0.613850
 7             sallow complexion            slight       0.603730


[ The diagnostic result ]

illness name   GanDamSpYeol   GanGiBuJog
                 0.998500      0.001183

Explanation :

     From the diagnostic process,the produced diagnostic  illness name
     is GanDamSpYeol. GanDamSpYeol is the causal illness name denoting
     the illness state caused by the high dampness and the high heat in
     the liver and gallbladder.
     Western style illness names appeared explicitly
     from the oriental causal illness name of GanDamSpYeol:
               Acute Hepatitis
               Hypertension
               Infectious disease
```

Figure 5: The diagnostic result of 'GanDamSpYeol' and its explanation : patient no.42

```
------------------------------------------------------------
The Medical Diagnostic result for the observed symptoms
------------------------------------------------------------

Sample 14

No.             Sympton              Linguistic Degree      value

 1                      jauntice      more serious       2.392400
 2                         fever      more serious       2.144000
 3                        nausea      more serious       2.480870
 4                  yellow urine      more serious       2.371230
 5          taut and rapid pulse           serious       1.600660
 6                     dizziness           serious       1.696530
 7             sallow complexion           serious       1.684460


[ The diagnostic result ]

* Probability of illness

illness name   GanDamSpYeol   GanGiBuJog
                 0.050552      0.987176

Explanation :

     From the diagnostic process,the produced diagnostic  illness name
     is GanGiBuJog. GanGiBuJog is the causal illness name denoting
     the illness state caused by the dificiency of liver-blood and the
     dificiency of liver-gi.
     Western style illness names appeared explicitly
     from the oriental causal illness name of GanGiBuJog:
               anemia
               hypotension
               chronic hepatitis
```

Figure 6: The diagnostic result of 'GanGiBuJog' and its explanation : patient no.14

medical diagnostic process is highly heuristic.

It is necessary to interpret the oriental disease name to the western style name for the easy communication because we are more familiar with the western style name in the modern times. This system has the explanation part that provide the western style disease name corresponded to the oriental causal illness state. Especially, in the oriental medical diagnosis, medical doctors regard the causal factors more important than the explicit illness name because they analyze the causal factors from the observed symptoms and give a treatment for curing the causes basically. They give a treatment with the gained causal illness name after diagnosis. But for helping the patients to understand well, the oriental medical doctors interpret the causal illness state to the illness names in the view of western style medical illness name and explain this to the patient.

5 Conclusion

The binary independent factor model based on the BYY harmony learning is suggested for automatic discovery medical diagnosis. We proposed a three layer factor model for the medical diagnosis. The model can not only recover the hidden factors from the visible observed symptoms but also produce the reasonable diagnostic results. Experiments on the binary hidden sources are demonstrated with success.

References

[1] Spearman,C., " General intelligence objectively determined and measured", *Am. J. Psycol. 15, 201-293,1904*.

[2] Xu, L., " BYY Harmony Learning, Independent State Space and Generalized APT Financial Analyses ", *IEEE Trans on Neural Networks*, in press, July, 2001.

[3] Xu, L., "Best Harmony, Unified RPCL and Automated Model Selection for Unsupervised and Supervised Learning on Gaussian Mixtures, Three-Layer Nets and ME-RBF-SVM Models", *International Journal of Neural Systems, Vol. 11*, No. 1, pp43-69, 2001.

[4] Xu, L., "Temporal BYY Learning for State Space Approach, Hidden Markov Model and Blind Source Separation", *IEEE Tr. Signal Processing 48*, 2132-2144, 2000.

[5] Xu, L., "Bayesian Kullback Ying-Yang Dependence Reduction Theory", *Neurocomputing 22*, No.1-3, 81-112, 1998.

[6] Xu, L.(1995&96), "A unified learning scheme: Bayesian-Kullback YING-YANG machine", *Advances in NIPS 8*, eds., D. S. Touretzky, et al, MIT Press, 1996, 444-450. Preliminarily on *Proc.ICONIP'95*, 1995, 977-988.