Proposal of Diagnostic Process Model for Computer based Diagnosis

Yasushi MATSUMURA\textsuperscript{a}, Toshihiro TAKEDA\textsuperscript{a},
Shiro MANABE\textsuperscript{a}, Hirokazu SAITO\textsuperscript{a}, Kei TERAMOTO\textsuperscript{a},
Shigeki KUWATA\textsuperscript{b} and Naoki MIHARA\textsuperscript{a}

\textsuperscript{a} Medical Informatics, Osaka University Graduate School of Medicine, Osaka, Japan

\textsuperscript{b} Division of Medical Informatics, National Cerebral and Cardiovascular Center
In spite of the improvements of diagnostic technology

**Diagnosis is still difficult**

For making diagnosis, knowledge about the relations between clinical findings and diseases is necessary.

These knowledge is in the memory of each medical doctor.

Part of it is lacking, part is ambiguous, part is invalid

**Diagnosis support system would be helpful**
Many types of diagnosis support system had been developed since the 1970s.

\[\text{e.g. INTERNIST-1} \Rightarrow \text{QMR}\]

Prominent system,
Encompass a broad range of internal medicine

In spite of their accuracy, most of them ceased to be used in actual clinical sites

\[\text{At that time, medical doctors did not use computer in clinical practice.} \Rightarrow \text{data entry to the system was a barrier to its use}\]

\[\text{these systems were difficult to sustain}
\text{because they depended on medical specialists to provide necessary knowledge.}\]
Introduction rate of Electronic Medical Record System in Japanese hospital with more than 400 beds

- **Physician Order Entry System**
- **Electronic Medical Record System**
Physician directly enter patients’ symptoms, physical findings, etc. into EMR
They sometimes use input template
The aim of our project

To Develop a diagnosis support system which links with EMR

First Step
Diagnosis support system which makes a list of possible causative diseases based on symptoms of a patient

Second Step
Include physical finding, and routine test results

Third Step
Include findings from special examinations
Topics of Presentation

1. Propose a diagnostic process model
   - Knowledge structure
   - Process of knowledge

2. Introduce a knowledge accumulating method for diagnosis support system
   makes a list of possible causative diseases based on symptoms of a patient

Lay weight on sustainability of knowledge base
   - Process the collection of small pieces of knowledge gleaned from textbooks or papers.
   - Non-specialists can submit knowledge into the system
   - A number of workers can submit knowledge concurrently
Diagnostic process under the case base would exist

### Case Base

<table>
<thead>
<tr>
<th>Disease</th>
<th>CF 1,1</th>
<th>CF 1,2</th>
<th>CF 1,3</th>
<th>CF 1,4</th>
<th>CF 1,5</th>
<th>CF 1,6</th>
<th>...</th>
<th>CF 1,N</th>
<th>CF 2,1</th>
<th>CF 2,2</th>
<th>CF 3,1</th>
<th>CF 3,2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disease1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>...</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Disease2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>...</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Disease3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>...</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

### Domain 1

### Domain 2

#### Domain 2

<table>
<thead>
<tr>
<th>CF 3,1</th>
<th>CF 3,2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

#### Patient Data

<table>
<thead>
<tr>
<th>CF 1,1</th>
<th>CF 1,2</th>
<th>CF 1,3</th>
<th>CF 1,4</th>
<th>CF 1,5</th>
<th>CF 1,6</th>
<th>...</th>
<th>CF 1,N</th>
<th>CF 3,1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>...</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
Calculation of disease probability from Case base

Frequency of disease i : \( F_i \)

the rate of cases which have the CF pattern A in disease i : \( R_{i,a} \)

The probability that the patient who has CF pattern A is suffering from disease d : \( Q_{d,a} \)

using Bayes' theorem

\[
Q_{d,a} = \frac{F_d R_{d,a}}{\sum_{i=1}^{N} F_i R_{i,a}}
\]
Actually a perfect case base does not exist.

In textbooks or papers a patient with a certain disease often has symptom S1 and S2, and sometimes S3.

**Calculation of \( R_{i,a} \) from occurrence rate of CFs:**

The occurrence rate of CF\( j \) in disease \( i \) : \( P_{i,j} \)

when CF\( j \) is found in a patient: \( x_j = 1 \)

when it is not: \( x_j = 0 \)

\[
R_{i,a} = \prod_{j=1}^{N} \left( P_{i,j} \times x_j + (1 - P_{i,j}) \times (1 - x_j) \right)
\]

under the assumption that these symptoms are developed independently in a disease.
Knowledge Structure

• Frequency of diseases in each category of sex and age group
  - Male, Female
  - infant term, child term, youth adulthood, middle-aged term and old age

• CFs in each disease and their occurrence rate in the disease

⇒ CF-disease relation

<table>
<thead>
<tr>
<th>Disease</th>
<th>CF</th>
<th>CF occurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subarachnoid bleeding</td>
<td>Headache acute occurrence</td>
<td>0.9</td>
</tr>
<tr>
<td>Subarachnoid bleeding</td>
<td>disturbance of consciousness</td>
<td>0.4</td>
</tr>
</tbody>
</table>
Match of different concept levels

Knowledge base (disease-CF relation)

Encephalitis  Headache  0.9
   Encephalitis   Acute onset headache   0.09 (0.9x0.1)
   Encephalitis   Chronic repetitive headache  0.63 (0.9x0.7)
   Encephalitis   Continuous headache  0.18 (0.9x0.2)

Patient’s symptom: Acute onset headache

naturally occurrence frequency.

Headache
   SubClassOf
   Acute onset headache  10%
       SubClassOf
       Chronic repetitive headache  70%
   SubClassOf
   Continuous headache  20%
Knowledge accumulating method for diagnosis support system that make a list of possible causative diseases based on symptoms

1. We made a thesaurus of symptoms and their attributes. They fall into 171 categories.

2. Symptom-disease relations are picked up from a textbook about pathognomy and submitted into a system. This task was carried out by 8 workers concurrently.

   At this time, 2,366 symptom-disease relations in 582 diseases were picked up.

3. The relative frequencies of diseases in each category of sex and age group were estimated from the disease data entered to EMR of Osaka University Hospital.

By entering symptoms of a patient, this system makes a list of possible causative diseases in probable order.
Collecting method for symptom-disease relations

1. Make a thesaurus of symptoms
2. Submit symptom-disease relations using this thesaurus
3. Categorize a symptom which is not included in the thesaurus into “other group”
4. Remake a thesaurus of symptoms that includes symptoms in “other group”
5. Convert symptom-disease relations according to the new thesaurus
Problems and solutions

Term of symptom and concept granularity are difficult to fix

- Continue to discuss this problem until a consensus is found.

Using the textbook which we had, we could not find the occurrence rate of a symptom in a disease

- Search for the pertinent information in other textbooks and papers.

Relative frequency of diseases in a university hospital is biased to rare diseases. It is not universal.

- Collect data from other hospitals and sources to correct the above bias.