**Introduction**

Our Approach in healthcare business
- Nihon Unisys is promoting information and communication technology (ICT) development based on business results of information platform construction related to medical care and health.

Our Motivation
- Our next step in business is to secondarily and beneficially utilize the data which gather from healthcare facilities. Our Motivation is to research and develop basic techniques to information extraction in medical fields, and review the availability to our medical information system solutions.

**Methods**

Dataset and Task Image

*Input Data Sample*

6月2日頃から腹痛が生じるとともに、嘔気・嘔吐出現

*Output Data Sample*

<s id="R104">2</s>日頃から腹痛が生じると、嘔気・嘔吐出現

Task 1) Extract complaint and diagnosis, and it’s modality attributes (modality="negation", "suspicion", "family")

Task 2) Give ICD-10 code on complaint and diagnosis

Method design in Task1 (Figure.1)
- Morphological Analysis by Mecab
- Discrimination by CRFs model
- Dictionary Perfect Matching

![Gather Data](Image)
![View/Update Data](Image)

Figure.1

- Test Data
- Train Data
- IOB2 Tag
- CRFs model
- Morphological Analysis by Mecab
- Discrimination by CRFs model
- Dictionary Perfect Matching
- Classified table of ICD-10
- Output Data
- ComeJoyo V5-1

Method design in Task2 (Figure.2)
- Replacement when
  - Perfect Matching
  - Partial Matching

Using three dictionary of
- Train data
- Standard Disease Master
- ICD-10

![Train Data](Image)
![Data Accuracy (%)](Image)

Table.1

<table>
<thead>
<tr>
<th>IOB2 Tag</th>
<th>Precision</th>
<th>Recall</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>88.14</td>
<td>74.53</td>
<td>74.53</td>
</tr>
<tr>
<td>Condition(Only)</td>
<td>81.83</td>
<td>69.38</td>
<td>75.09</td>
</tr>
<tr>
<td>Condition(Positive)</td>
<td>70.17</td>
<td>56.34</td>
<td>62.50</td>
</tr>
<tr>
<td>Condition(Negative)</td>
<td>51.05</td>
<td>54.61</td>
<td>52.77</td>
</tr>
<tr>
<td>Condition(Suspicion)</td>
<td>45.45</td>
<td>12.20</td>
<td>19.23</td>
</tr>
<tr>
<td>Condition(Family)</td>
<td>66.67</td>
<td>53.85</td>
<td>59.57</td>
</tr>
</tbody>
</table>

Table.2

<table>
<thead>
<tr>
<th>Data</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GoldStandard+ICD</td>
<td>69.40</td>
</tr>
</tbody>
</table>

**Results**

**Task1 (Table.1)**
- We could extracted time and condition(only) related expressions in a high accuracy. On the other hand it was difficult to extract the modality types, especially “Suspicion”.

**Task2 (Table.2)**
- The result of Task2 is shown in Table 2. We used “GoldStandard+ICD” as test data which was distributed by MedNLP-2 organizer as correct answer data of Task1.

**Discussions**

Modality Extraction in Task1
- It was difficult to extract the modality types in a high accuracy. We think that adding more rules to complement our algorithm is needed.

Consideration of Local recording rule in each hospitals
- It is important to consider the local rule of each hospitals and add these rules to the algorithm to use widely as a function of medical information system, because each hospital have each rule of recording document.

For beneficial Data Mining
- It is very important not only to develop these supporting system but also to consider a purpose to analysis data which are extracted from medical data because the requirements of extracting data depend strongly on the purposes of data analysis. We think that each approaches are needed.

**Bibliography**


...