Introduction of Medical NLP for Web Document (MedWeb)

The 4th Round of MedNLP Task

Aramaki E, Wakamiya S, Kano Y, Ohkuma T, Morita M

1NAIST, 2Shizuoka Univ, 3Fuji Xerox, 4Okayama Univ
MedNLP Milestones

**Medical Documents**

“ACTOS was prescribed at St Luke’s Hospital, and his leg was swollen.”

**De-identification**

“ACTOS was prescribed at [redacted] Hospital, and his leg was swollen.”

**NER (Named Entity Recognition)**

“Pioglitazone was prescribed at [redacted] Hospital, and his leg was edema.”

**Decision Support**

“$\Psi(\Theta) \cdot (\Theta) \Psi > \& \# \& \Theta : $※$\Theta$?”

**Better Clinical Solutions**
We’ve done everything?
From In-house data to Open web data

NO!!
Massive texts are available outside of hospitals
We’ve done everything?
From In-house data to Open Web data

1: Twitter
2: Blog

Medical Documents
"ACTOS was prescribed at St Luke’s Hospital, and his leg was swollen."

De-identification
"ACTOS was prescribed at Hospital, and his leg was swollen."

NER (Named Entity Recognition)
"Pepitazone was prescribed at Hospital, and his leg was edema."

Better Clinical Solutions
1,000 tweets × 14 symptoms × 3 languages

- Collecting tweets by a crowdsourcing
- Translating Japanese tweets into English and Chinese

<table>
<thead>
<tr>
<th>Category</th>
<th>Tweet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neg</td>
<td>@□□□ 大丈夫か？全身の関節がいたいなら、インフルエンザかもしれないよ</td>
</tr>
<tr>
<td>Neg</td>
<td>@□□□ 時間があれば検査だけ受けておいたほうがいいかもしれない．w38度越え始めたら、一旦熱引いたとしてもインフルエンザの可能性はありますので．</td>
</tr>
<tr>
<td>Pos</td>
<td>私インフルエンザかもしれない． Tops</td>
</tr>
<tr>
<td>Pos</td>
<td>昨日早退した人はインフルエンザだった、体調わるい人たくさん</td>
</tr>
<tr>
<td>Pos</td>
<td>@□□□ (     ぞ） Œーインフルエンザでしたか…お大事にです…たいぶ症状がでてこないとお仕事休めないというのも大変ですね…</td>
</tr>
<tr>
<td>Pos</td>
<td>@□□□ わたし、インフルエンザ、なう．指定のインスト行ける♡</td>
</tr>
</tbody>
</table>

**TASK:** To classify a given tweet into patient or not

(Pos) (Neg)
Annotation guideline for 14 symptoms (only Japanese)

- influenza 「インフル」
- sick 「風邪」
- sore throat 「喉の痛み」
- chill 「寒気」
- runny nose 「鼻水・鼻づまり」
- cough 「咳・痰」
- fever 「熱」
- headache 「頭痛」
- rash 「発疹」
- vomiting 「悪心」
- spasm 「痙攣」
- diarrhea 「下痢」
- stomachache 「腹痛」
- asthma 「喘息」

Designed based on the advises by National Institute of Infectious Diseases (NIID)
(2) Blog Track

Blog contains various subjective information on diseases

110 Japanese articles written by actual patients (mostly cancer patients)

TASK: To extract complaints, drug names and disease names
Potential Applications

Web-based infectious disease surveillance/warning system

Communication support between patients and doctors
From In-house data to Open Web data

MedWeb

http://mednlp.jp/medweb/NTCIR-13/
E-mail: medweb@is.naist.jp

supported by

NIID 国立感染症研究所

Mediaid

AMED