Hidden concept relation discovery in Biomedical Literature using co-occurrence

Yan Huang
Problem Definition

Concept A

Link

Concept

Concept B
## Method

**Open Discovery**

### Concept A

<table>
<thead>
<tr>
<th>Link Concept</th>
<th>Co-occurrence count</th>
</tr>
</thead>
<tbody>
<tr>
<td>B1</td>
<td>Bc1</td>
</tr>
<tr>
<td>B2</td>
<td>Bc2</td>
</tr>
<tr>
<td>B3</td>
<td>Bc3</td>
</tr>
<tr>
<td>......</td>
<td>......</td>
</tr>
<tr>
<td>Bn</td>
<td>Bcn</td>
</tr>
</tbody>
</table>

### Target Concept

<table>
<thead>
<tr>
<th>Target Concept</th>
<th>Co-occurrence count</th>
</tr>
</thead>
<tbody>
<tr>
<td>C11</td>
<td>Cc11</td>
</tr>
<tr>
<td>C12</td>
<td>Cc12</td>
</tr>
<tr>
<td>C13</td>
<td>Cc13</td>
</tr>
<tr>
<td>......</td>
<td>......</td>
</tr>
<tr>
<td>C1k</td>
<td>Cc1k</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Target Concept</th>
<th>Co-occurrence count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cn1</td>
<td>Ccn1</td>
</tr>
<tr>
<td>Cn2</td>
<td>Ccn2</td>
</tr>
<tr>
<td>Cn3</td>
<td>Ccn3</td>
</tr>
<tr>
<td>......</td>
<td>......</td>
</tr>
<tr>
<td>Cnk</td>
<td>Ccnk</td>
</tr>
</tbody>
</table>

---

Data Set

Unified Medical Language System (UMLS)

Medical Subject Headings (MeSH)

MEDLINE co-occurrences (MRCOC)

CUI1

Starred Frequency

D000001|C0000699|D000515|C0002191|1|0|1989|RST|0|0|1|ZN|0|0|0|0
D000001|C0000699|D000515|C0002191|1|0|1990|RST|0|0|1|ZN|1|0|0|0
D000001|C0000699|D000515|C0002191|1|0|1993|RST|0|0|1|ZN|0|0|0|0
D000001|C0000699|D000515|C0002191|1|1|1995|RST|0|0|1|ZY|0|0|0|0
D000001|C0000699|D000515|C0002191|2|0|2004|MBD|0|0|2|ZN|0|2|0|0
D000001|C0000699|D000515|C0002191|1|0|2011|MED|0|0|1|ZN|0|0|0|0
Work Flow

Raw MRCOC File → Useful data → Frequent Patterns

Starred Frequency >= 5
(Support)

One Seed
Open Discovery → Compute Confidence → Top Candidates

Two Seeds
Closed Discovery → Compute Confidence → Related or not
Problem and Future work

• Hard to evaluate
  use data before 1980, see if the results have relation in the literature after 1980.

• SVM?
  Features: All co-occurrence, starred co-occurrence, co-occurrence in one sentence......
  Labeled data

• Real-time Queries?
MapReduce
MapReduce

1:2,3,4
2:156
3:1,5
4:1,5
5:2,3,4
6:2
Thank you!

Q&A