Semantic IR: Exploring Dependency and Word Embedding Features in Biomedical IR

Rastegar-Mojarrad, M., Komandur Elayavilli, R., Liu, S., Wang, Y., Shen, F., Liu, H.
BELTracker

• Task 2: Retrieving evidence sentence for a given BEL
  • Identifying at most 10 evidence sentences

• Two architectures
  • 2015 and 2017 competitions
  • Co-occurrence based
  • IE-based

• Main components
  • Retrieval component
  • Ranking component
Co-occurrence based (2015)

- Indexing abstract and full-text
  - Elastic search

- Retrieval component
  - Translating the given BEL statement into a query
    - Identifying elements
    - Finding synonyms
      - Existing resources & expert-generated dictionaries
  - Retrieving relevant documents
BEL statement

\[ p(\text{MGI:tnf}) \text{ increases } p(\text{MGI:creb1}, \text{pmod}(P,S,133)) \]

Identifying elements

Namespace: Entity
1) MGI:tnf
2) MGI:creb1

Relation
increases

Function
1) p
2) pmod

Argument
1) P
2) S

Finding synonyms

- tnf: tnf alpha, tnfα, ...
- increases: induce, activate, ...
- pmod: phosphorylation, ...
- etc...

Query

Mandatory elements
- tnf or synonyms
- creb1 or synonyms

Optional elements
- increases or synonyms
- pmod or synonyms
- P or synonyms

Querying abstracts and full-texts
Co-occurrence based

• Ranking component

• Top 1000 results
  • Splitting into sentences
  • Removing non-relevant sentences

• Ranking sentences based on occurrence of elements
  • Manual weight assignment for elements
IE-based (2017)

• Indexing only useful sentences
  • Using Semantic Medline

• Training classifiers for ranking
  • For each element

• Semantic
  • Word embedding
  • Dependency embedding
Semantic Medline

• **SemRep**
  - Extracts semantic predications from biomedical literature
  - Biomedical entities
  - ~30 predicates
    - treat, affect, interacts with, …, *coexist with*

• **Semantic Medline Database**
  - A relational database
  - ~ 200 million sentences
  - ~ 90 million predications
Ranking component

• Using the same retrieval component

• Training classifier for ranking
  • Pair classification (BEL, sentence)
  • Positives from the training set
  • How to define negative instances ???
    • Randomly selecting a sentence for a BEL
      • Masking entity names
      • Not the same relation (inc. or dec. )
      • Not the same function

• Train three classifiers
  • Distant supervision
First classifier

• Existence of relationship
  • \( \text{cat}(	ext{HGNC:SPN}) \) increases \( \text{gtp}(	ext{HGNC:RAP1A}) \)
  • *RasGRP2 also catalyzes nucleotide exchange on* \( \text{Rap1a} \) *and* \( \text{SPN} \), *but this RapGEF activity is less potent than that associated with* \( \text{CalDAG-GEFI} \).

• Instances (Pair of BEL statement and sentence)
  • Negative instances from Semantic medline
    • Entities: the same as randomly selected BEL
    • Predicate: *coexist with*

• Features
  • Unigram, Bi-grams, Word-embedding
Second classifier

• Function

<table>
<thead>
<tr>
<th>Subject</th>
<th>Predicate</th>
<th>Object</th>
</tr>
</thead>
<tbody>
<tr>
<td>cat(HGNC:RASGRP2)</td>
<td>increases</td>
<td>gtp(HGNC:RAP1A)</td>
</tr>
</tbody>
</table>

• **RasGRP2** also **catalyzes** nucleotide exchange on Rap1, but this RapGEF activity is less potent than that associated with CalDAG-GEFI.

• One classifier for each function

• Instances: BEL term+ sentence
  • Positive instances from the training data
  • Negative instances others functions
Second classifier

• Features
  • Surrounding terms (window of 3-5)
  • Dependency embedding

RasGRP2 also catalyzes nucleotide exchange on Rap1, but this RapGEF activity is less potent than that associated with CalDAG-GEFI.
Third classifier & final score

- Relation
  - cat(HGNC:SPN) increases r(HGNC:CD40LG)

- Two classes: increase and decrease
  - Instances: BEL and sentence
  - No challenge in identifying instances for each class
  - Fea. : Unigrams, bigrams, expert-generated dictionary, ..

- Final score for each retrieved sentence
  \[ 0.40 \times P_{c1} + 0.50 \times P_{c2} + 0.10 \times P_{c3} \]
Data

• Training set
  • Pair of BEL and corresponding sentence
  • 11066 BEL statements extracted from 6354 sentences
  • No entity or relation annotation

• Test set
  • 100 BEL statement for each competition
  • Manually evaluated by expert
Evaluation

- Precision
  - \( \frac{TP}{(FP + TP)} \)

- Criteria
  - **Full**: The sentence contains the complete BEL statement.
  - **Relaxed**: Has necessary context and/or biological background to enable extraction of BEL statement.
  - **Context**: Even though the complete or partial BEL statement cannot be extracted from the sentence, it provides the necessary context for the BEL statement.
Evaluation

• Mean Average Precision (MAP)
  • For each query
    • Precisions for relevant docs considering their rank
    • Ave. the precisions : AP
  • Mean of all APs for all Qs

• Three ranking scenarios
  • Worst: All the TPs are ranked after all FPs
  • Random: Randomly reordered the results 2000 times
  • Best: All TPs are ranked before all FPs
Results (2015 test data)

• Precision (co-occurrence based)

<table>
<thead>
<tr>
<th>Criteria</th>
<th>True positive</th>
<th>False positive</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full</td>
<td>316</td>
<td>490</td>
<td>39.20</td>
</tr>
<tr>
<td>Relaxed</td>
<td>429</td>
<td>377</td>
<td>53.22</td>
</tr>
<tr>
<td>Context</td>
<td>496</td>
<td>310</td>
<td>61.53</td>
</tr>
</tbody>
</table>

• MAP

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Worst</th>
<th>Random</th>
<th>Co-occ.</th>
<th>IE-based</th>
<th>Best</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full</td>
<td>31.7</td>
<td>46.5</td>
<td>49.0</td>
<td>56.96</td>
<td>74.2</td>
</tr>
<tr>
<td>Relaxed</td>
<td>45.9</td>
<td>58.4</td>
<td>62.1</td>
<td>65.05</td>
<td>80.4</td>
</tr>
<tr>
<td>Context</td>
<td>55.2</td>
<td>65.7</td>
<td>68.9</td>
<td>73.15</td>
<td>83.5</td>
</tr>
</tbody>
</table>
Results (2017 test data)

- Two runs (top 5 evidence sentences)

<table>
<thead>
<tr>
<th>Run</th>
<th>Criteria</th>
<th>True positive</th>
<th>False positive</th>
<th>Precision</th>
<th>MAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>ElasticSearch</td>
<td>Full</td>
<td>117</td>
<td>265</td>
<td>30.6</td>
<td>59.6 (+9.4)</td>
</tr>
<tr>
<td>Our system</td>
<td>Full</td>
<td>121</td>
<td>261</td>
<td>31.7 (+1.1)</td>
<td>50.2</td>
</tr>
<tr>
<td>ElasticSearch</td>
<td>Partial</td>
<td>175</td>
<td>207</td>
<td>45.8</td>
<td>77.5 (+0.8)</td>
</tr>
<tr>
<td>Our system</td>
<td>Partial</td>
<td>192</td>
<td>190</td>
<td>50.3 (+4.5)</td>
<td>76.7</td>
</tr>
</tbody>
</table>
Discussion

• Solved issues
  • Response time
  • Extensively relied upon the lexical feature
    • Without any consideration of semantics

• Dictionaries
  • 16% of BEL entities not in the corresponding sentences
  • No results for 7 BELs
    • Focusing on Full retrieval not Partial sentences
Discussion

• Only sentences in abstracts

• Ranking
  • Negative instances for the first classifier
  • No annotation for the mentions

• Evaluation
  • Need for an expert
Thank you
Classifiers result

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Feature sets</th>
<th>Model</th>
<th>F-measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entity</td>
<td>Unigram, bi-gram, Embedding of terms between entities</td>
<td>SVM</td>
<td>0.946</td>
</tr>
<tr>
<td>Function</td>
<td>Window of 3-5, unigram, bigrams, Wo. Emb, Dep. Emb</td>
<td>SVM</td>
<td>0.821</td>
</tr>
<tr>
<td>Relation</td>
<td>Unigram, bi-grams, …</td>
<td>SVM</td>
<td>0.835</td>
</tr>
</tbody>
</table>