IBM Research

Expanding Query Answers on Medical Knowledge Bases

Chuan Lei
Vasilis Efthymiou
Rebecca Geis
Fatma Özcan
Querying medical knowledge bases
Query relaxation

Problem:
Users do not always formulate their queries precisely to match the terms in the KB
➢ No answer or incomplete answers returned

Goal:
Query relaxation (QR) transforms the query in a way that the user’s intent is better represented
➢ greatly improving the flexibility and usability of a medical KB

Contributions:
• an effective offline external knowledge source incorporation
• a novel similarity metric to identify semantically related concepts
• a programmatic way to incorporate our QR into existing systems
• experimental evaluation shows our QR outperforms existing methods
Two-phase approach (overview)

**Offline phase (aka external knowledge source incorporation):**
(i) Initialize the set of contexts, (ii) compute concept frequencies, (iii) generate mappings

**Online phase (aka online query relaxation):**
(i) map query term to external concept, (ii) return top-k external concepts
External knowledge source incorporation

Mapping medical KB to external knowledge source
➢ exact match / fuzzy match / embeddings / …

The context of a query term can be represented by a relationship and its associated concepts from the domain ontology

Concept frequency

\[ freq(A) = |A| + \sum_{A_i \in A} freq(A_i) \]

Information content-based similarity

\[ IC(A) = -\log(freq(A)) \]

\[ sim_{IC}(A, B) = \frac{2 \times IC(lcs(A, B))}{IC(A) + IC(B)} \]
Online query relaxation

**Generalization vs specialization**

The weight of a path connecting two external concepts A and B: 

\[ p_{A,B} = \prod_{i}^{D} w_i^{D-i} \]

Overall concept similarity: 

\[ \text{sim}(A,B) = p_{A,B} \times \text{sim}_{IC}(A,B) \]
Putting it all together

• Given a query term $q$, the query relaxation method
  1. finds an external concept $A$ that matches $q$
  2. searches for the external concepts within $r$ distance from $A$
  3. retrieves the top-$k$ pre-computed similarity between $A$ and each external concept in its neighborhood. Top-$k$ relaxed results are returned based on their overall similarity scores

• $r$ can be:
  – set as a fixed value by empirical studies, or
  – dynamically decided if a fixed $r$ cannot provide $k$ results

• $k$ can be application-specific or defined by users
Integration with IBM Watson Assistant

Experimental evaluation

Accuracy of mapping methods

<table>
<thead>
<tr>
<th>Methods</th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXACT</td>
<td>100</td>
<td>83.33</td>
<td>90.01</td>
</tr>
<tr>
<td>EDIT</td>
<td>96.36</td>
<td>88.33</td>
<td>92.17</td>
</tr>
<tr>
<td>EMBEDDING</td>
<td>96.49</td>
<td>91.67</td>
<td>94.02</td>
</tr>
</tbody>
</table>

Overall effectiveness of query relaxation (QR)

<table>
<thead>
<tr>
<th>Methods</th>
<th>P@10</th>
<th>R@10</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>QR</td>
<td>90.51</td>
<td>82.64</td>
<td>86.40</td>
</tr>
<tr>
<td>QR-no-context</td>
<td>85.45</td>
<td>77.27</td>
<td>81.15</td>
</tr>
<tr>
<td>QR-no-corpus</td>
<td>78.23</td>
<td>70.91</td>
<td>74.39</td>
</tr>
<tr>
<td>IC</td>
<td>75.55</td>
<td>68.18</td>
<td>71.68</td>
</tr>
<tr>
<td>Embedding-pre-trained</td>
<td>66.14</td>
<td>60.13</td>
<td>62.99</td>
</tr>
<tr>
<td>Embedding-trained</td>
<td>79.37</td>
<td>71.81</td>
<td>75.40</td>
</tr>
</tbody>
</table>

Setup
- KB: IBM Micromedex
- External knowledge source: SNOMED CT
- Corpus: a few thousand in-depth documents describing drugs, findings, adverse effects

Results
- IC baseline is not as good as QR even the variations without context or corpus information
- QR without contextual information is reasonable
- QR without corpus is much worse
- pre-trained* is off-the-shelf, but worst results
- trained: using glove and fasttext

* http://bio.nlplab.org
Experimental evaluation – user study

User study with 20 medical SMEs:
Watson Assistant with and without query relaxation (QR)

<table>
<thead>
<tr>
<th>Score</th>
<th>QR</th>
<th>no QR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T1</td>
<td>T2</td>
</tr>
<tr>
<td>1 (Very dissatisfied)</td>
<td>2.1%</td>
<td>10.55%</td>
</tr>
<tr>
<td>2 (Dissatisfied)</td>
<td>10.35%</td>
<td>11.07%</td>
</tr>
<tr>
<td>3 (Okay)</td>
<td>25.59%</td>
<td>29.33%</td>
</tr>
<tr>
<td>4 (Satisfied)</td>
<td>35.21%</td>
<td>33.37%</td>
</tr>
<tr>
<td>5 (Very satisfied)</td>
<td>26.85%</td>
<td>15.68%</td>
</tr>
<tr>
<td>AVG</td>
<td>3.73</td>
<td>3.31</td>
</tr>
</tbody>
</table>

T1: for 20 fixed concepts, SMEs pick 20 questions
T2: SMEs are free to ask 10 questions about anything

Observations
• QR improved the user experience in both tasks on average by 20% compared to no QR
• T1 results better than T2
• User feedback for not satisfying answers:
  – expected answers are not contained in the given KB
  – not ideal conversational flow (irrespective of QR results)
  – the amount of information returned is overwhelming
Summary

• A novel two-phase query relaxation method
  – leverages external knowledge sources
  – empowers semantically related concepts with a novel similarity metric

• Integration with two exemplary systems
  – a conversational system
  – a natural language query system

• Our method outperforms state-of-the-art ones in precision and recall
• User study shows our method
  – expands the query results
  – improves their quality for medical KBs
Thank you!