Entity-Supported Summarization of Biomedical Abstracts

Frederik Schulze and Mariana Neves

December 12th, 2016
Automatic Summarization in Biomedicine

- Automatic summarization of multiple texts can provide summaries specific to the user’s information needs.

---

**Automatic Summarization in Biomedicine**

- Automatic summarization of multiple texts can provide summaries specific to the user’s information needs.

---

**TP53**

- **Official Symbol**: TP53
- **Official Full Name**: tumor protein p53
- **Primary source**: HGNC
- **See related**: Ensembl:ENSG00000141510, HPRD:01859, MIM:191170, Vega:OTTHUMG00000162125
- **Gene type**: protein coding
- **RefSeq status**: REVIEWED
- **Organism**: Homo sapiens
- **Lineage**: Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo
- **Also known as**: P53; BCC7; LFS1; TRP53
- **Summary**: This gene encodes a tumor suppressor protein containing transcriptional activation, DNA binding, and oligomerization domains. The encoded protein responds to diverse cellular stresses to regulate expression of target genes, thereby inducing cell cycle arrest, apoptosis, senescence, DNA repair, or changes in metabolism. Mutations in this gene are associated with a variety of human cancers, including hereditary cancers such as Li-Fraumeni syndrome. Alternative splicing of this gene and the use of alternate promoters result in multiple transcript variants and isoforms. Additional isoforms have also been shown to result from the use of alternate translation initiation codons (PMIDs: 12032546, 20937277). [provided by RefSeq, Feb 2013]
- **Orthologs**: mouse all

Automatic Summarization in Biomedicine

- Automatic summarization of multiple texts can provide summaries specific to the user’s information needs.

What symptoms characterize the Muenke syndrome?

“There’s syndrome is an autosomal dominant disorder characterized by coronal suture craniosynostosis, hearing loss, developmental delay, carpal and tarsal fusions, and the presence of the Pro250Arg mutation in the FGFR3 gene. Muenke syndrome is characterized by coronal craniosynostosis (bilateral more often than unilateral), hearing loss, developmental delay, and carpal and/or tarsal bone coalition. Tarsal coalition is a distinct feature of Muenke syndrome and has been reported since the initial description of the disorder in the 1990s.”
Proposed methods

- **Extended LexRank**: extension of LexRank graph-based algorithm with information from named-entity recognition.

- **EntityRank**: similarity between the sentences to generate a sentence graph, but based on the named entities.
Proposed methods

- **EntityRank for question answering**: we changed EntityRank to provide a bonus score for sentences that contain words which overlap with the questions.

- **EntityRank for gene summaries**: we created a bonus score that reflects how suited a sentence is for belonging to a gene summary, based on the 15 most frequent biomedical terms, such as ”Proteins”, ”Genes”, ”protein location”, ”encoding”, ”variant”, ”family”, ”last name”, ”variant”, ”receptor”, ”receptor cells”, ”mutation”, ”numerous”, ”function”, ”enzymes”, and ”DNA”.

Evaluation (two datasets)

- 1009 questions from BioASQ challenge
- Entrez Gene summaries: 9555 summaries for training, 1974 for test
Results

- Comparison to LexRank (BioASQ dataset)
Results

- Evaluation of the bonus factor (BioASQ dataset)
Results

- Evaluation for gene summaries
<table>
<thead>
<tr>
<th></th>
<th>PHASE B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BATCH 1</td>
</tr>
<tr>
<td>EXACT</td>
<td>1st</td>
</tr>
<tr>
<td></td>
<td>(SELECTION STRATEGY: AVG RANKING IN THREE TYPES: Y/N-ACC, FACT-MRR, LIST MEAN F-MEASURE)</td>
</tr>
<tr>
<td>IDEAL</td>
<td>1st</td>
</tr>
<tr>
<td></td>
<td>(SELECTION STRATEGY: MEAN MANUAL SCORE)</td>
</tr>
</tbody>
</table>

(http://www.bioasq.org/participate/fourth-challenge-winners)
Thank you!

Please visit our poster!

Looking forward to answering your questions!