Text Analysis of Scientific Papers Using NooJ: a Case Study With Vitamin D Supplementation Debates

Lesia Kaigorodova
UIIP, National Academy of Sciences of Belarus
Minsk, Belarus
Automatic Opinion Analysis / Opinion Mining

Problem

- Opinion mining is an approach to natural language processing (NLP) that identifies the emotional tone behind a body of text.

- A popular way to determine and categorize opinions about a product, service or idea.
Vitamin D Supplementation Debate. Problem Statement

This work is dedicated to text mining of scientific papers with NooJ regarding vitamin D consumption, probably the most popular topic nowadays in the era of supplements.

We will try to design a system that would provide answers to such questions as:

- do we benefit from vitamin D intake or is it harmful?
- can we provide medical practitioners with up-to-date knowledge in the domain?
Dataset Description

- **Source.** The data was scraped from the PubMed database of the NCBI (United States National Center for Biotechnology Information) website. It has the greatest selection of articles in the field of medical domain in English.

- **Size.** Overall metatdata of 83077 articles were downloaded from the PubMed database that refer to the keyword “vitamin D”.

- **Structure.** The data comprises *title, list of authors, date, journal, citation* and an *abstract*. Some abstracts contain more structured information in the form of sections such as “conclusion”, “objectives”, “method”, “results”, “design”, “setting”.
Problem Solution With NooJ

**Step 1.** Constructing look-up dictionaries for *positive*, *negative* and *neutral* terms

**Step 2.** Syntactic grammar to combine the terms

**Step 3.** Evaluation
Problem Solution With NooJ. Step 1

Constructing look-up dictionaries for **positive**, **negative** and **neutral** terms

- ‘Vitamin D’ term (<vitd>)
- ‘Supplementation’ term (<suppl>)
Problem Solution With NooJ. Step 1

Constructing look-up dictionaries for **positive(<benefit>)**, **negative(<harm>)** and **neutral (<neutral>)** terms

<harm>

<benefit>

<neutral>
Problem Solution With NooJ. Step 1

Constructing look-up dictionary for:

- Negation
- Annihilation (minimization) of effects, conditions
Problem Solution With NooJ. Step 2
Problem Solution With NooJ. Step 2

Construction of grammars
Problem Solution With NooJ. Step 3

Evaluation

- Imbalanced dataset (~85% positive:~15% negative):
  
  the dominant opinion on vitamin D intake is positive in scientific papers!

- Neutral statements included. Ex:

  Effects of six months of vitamin D supplementation in patients with heart failure/SupplRecommend+Sentiment=Negative
  vitamin D supplementation and lifestyle advice of healthy/SupplRecommend+Sentiment=Positive
  vitamin D supplementation in the improvement/SupplRecommend+Sentiment=Positive

  CONCLUSIONS: The evidence addressing the use of vitamin D for chronic pain/SupplRecommend+Sentiment=Negative
  : a random sunlight e of critical now con
Problem Solution With NooJ. Step 3

Evaluation (from perspective of targeting negative labels)

- After filtering neutral statements:

<table>
<thead>
<tr>
<th></th>
<th>predicted</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>pos</td>
<td>neg</td>
<td></td>
</tr>
<tr>
<td>actual</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pos</td>
<td>867</td>
<td>43</td>
<td></td>
</tr>
<tr>
<td>neg</td>
<td>19</td>
<td>82</td>
<td></td>
</tr>
</tbody>
</table>

Precision = 0.66
Recall = 0.81
F-measure = 0.73
Further improvement

- add neutral tag/sentiment for statements
- add list of deceases and conditions to the <harm> effect
- enrich dictionaries with new words
Thank you!