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

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

'Vitamin D' term (<vitd>)

Dictionary contains 72 entries	
7	
vitd,N+Medic+Vitamin	
#D, vitd, N+Medic+Vitamin	
D2, vitd, N+Medic+Vitamin	
D3, vitd, N+Medic+Vitamin	
VD, vitd, N+Medic+Vitamin	
VD2, vitd, N+Medic+Vitamin+UNAMB	
VD3, vitd, N+Medic+Vitamin+UNAMB	
<pre>#vitD, vitd, N+Medic+Vitamin</pre>	
vitD2, vitd, N+Medic+Vitamin+UNAMB	
vitD3,vitd,N+Medic+Vitamin+UNAMB	
vitamin D, vitd, N+Medic+Vitamin+UNAMB	
vitamin D2, vitd, N+Medic+Vitamin+UNAMB	
vitamin D3, vitd, N+Medic+Vitamin+UNAMB	
cholicalciferol, vitd, N+Medic+Vitamin	
cholecalciferol, vitd, N+Medic+Vitamin	
"1,25-dihydroxyvitamin", vitd, N+Medic+Vitamin+UNF	MB
"1,25-hydroxyvitamin", vitd, N+Medic+Vitamin+UNAME	3
"25-hydroxyvitamin", vitd, N+Medic+Vitamin+UNAMB	
"25-dihydroxyvitamin", vitd, N+Medic+Vitamin+UNAME	3
dihydroxyvitamin, vitd, N+Medic+Vitamin	
hydroxyvitamin, vitd, N+Medic+Vitamin	
"1,25-dihydroxyvitamin D",vitd,N+Medic+Vitamin+U	JNAMB
"1,25-hydroxyvitamin D", vitd, N+Medic+Vitamin+UNA	MB
"25-hydroxyvitamin D", vitd, N+Medic+Vitamin+UNAME	3
"25-dihydroxyvitamin D", vitd, N+Medic+Vitamin+UNA	MB
dihydroxyvitamin D.vitd.N+Medic+Vitamin+UNAMB	1. FASTER 199

'Supplementation' term (<suppl>)

	🖳 [Modified] C:\Users\kaigorodov\Documents\NooJ\en\Lexical Analysis\suppl.dic			
# # #	Special Characters: '\' '"' ' ' ',' '+' '-' '#'			
sı	uppl,N+Medic			
sn in a c t : u : u : n : r : p : a	upplement, suppl, N+Medic upplementation, suppl, N+Medic ntake, suppl, N+Medic didtive, suppl, N+Medic onsumption, suppl, N+Medic reatment, suppl, N+Medic se, suppl, N+Medic se of, suppl, N+Medic sing, suppl, N+Medic egulation, suppl, N+Medic rescription, suppl, N+Medic doption, suppl, N+Medic			

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

<harm>



<benefit>

[Modified] C:\Users\kaigorodov\Documents\NooJ\en\Lexi.	• •
#uerbrur	^
benefit,N+Benefit	
improvement, benefit, N+Benefit	
help,benefit,N+Benefit	
progress, benefit, N+Benefit	
satisfaction, benefit, N+Benefit	
favor, benefit, N+Benefit	
favour, benefit, N+Benefit	
aid, benefit, N+Benefit	
assistance,benefit,N+Benefit	
relief, benefit, N+Benefit	
support, benefit, N+Benefit	
good, benefit, N+Benefit •	
success, benefit, N+Benefit	
protection, benefit, N+Benefit	
benefitial, benefit, A+Benefit	
positive, benefit, A+Benefit	
good, benefit, A+Benefit	
promising, benefit, A+Benefit	
helpful, benefit, A+Benefit	
healthy, benefit, A+Benefit	
successful, benefit, A+Benefit	
protectful, benefit, A+Benefit	

<neutral>

#neutral	
neutral,A+Neutral	
result, neutral, N+Neutral	
development, neutral, N+Neutral	
outcome, neutral, N+Neutral	
feedback, neutral, N+Neutral	
effect, neutral, N+Neutral	
impact, neutral, N+Neutral	
consecuence neutral N+Neutral	
association, neutral, N+Neutral	
response, neutral, N+Neutral	
trigger, neutral, N+Neutral	
state, neutral, N+Neutral	
phase, neutral, N+Neutral	
condition, neutral, N+Neutral	

Constructing look-up dictionary for:

- Negation
- Annihilation (minimization) of effects, conditions



1	🖳 C:\Users\kaigorodov\Documents\NooJ\e 💼 🔳 🎫
	Dictionary contains 70 entries
	barely, negative, ADV+Negativ
	hardly, negative, ADV+Negativ
	lower, negative, V+Negativ
	affect, negative, V+Negativ
	adverse, negative, V+Negativ
	reverse, negative, V+Negativ
	decrease, negative, V+Negativ
	retain, negative, V+Negativ
	minimize, negative, V+Negativ
	fail, negative, V+Negativ
	ruin, negative, V+Negativ •
	attenuate, negative, V+Negativ
	alleviate, negative, V+Negativ
l	collapse, negative, V+Negativ
	falsify, negative, V+Negativ
	withhold, negative, V+Negativ
	slow, negative, V+Negativ
	avoid, negative, V+Negativ
	exit, negative, V+Negativ
	quit, negative, V+Negativ
	alleviate, negative, V+Negativ
	reduce, negative, V+Negativ
	alleviate, negative, V+Negativ
l	prevent, negative, V+Negativ
1	stop, negative, V+Negativ
1	annihilate, negative, V+Negativ
1	delay, negative, V+Negativ





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=""></supplrecommend+sentiment=>	: a randor
D deficiency. Targeted strategies including	vitamin D supplementation and lifestyle advice of healthy/ <supplrecommend+sentiment=positive></supplrecommend+sentiment=positive>	sunlight e
answer regarding the role of	vitamin D supplementation in the improvement/ <supplrecommend+sentiment=positive></supplrecommend+sentiment=positive>	of critical
CONCLUSIONS: The evidence addressing the	use of vitamin D for chronic pain/ <supplrecommend+sentiment= negative=""></supplrecommend+sentiment=>	now con

Evaluation (from perspective of targetting negative labels)

• After filtering neutral statements :

		predicted	
		pos	neg
actual	pos	867	43
	neg	19	82

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!