Adaptation of IDPT system based on patient-authored text data using NLP: Supplement resources

4 maj 2020

Supplement Documents

1 Introduction

This document is created as the supplement resources for the paper entitled `Adaptation of` `IDPT system based on patient-authored text data using NLP: Supplement resources` by Mukhiya et. al. (`itsmeskm99@gmail.com`).

2 Prepossessing algorithm

3 PHQ-9 questionnaire symptoms

4 Seed term generation algorithm

Algorithm 1: Preprocessing test text sentences

---
input : A patient authored test `text`
output : Processed test text

---
1   text ← encode(`text`, `UTF-8`);

2   text ← lower(`text`);

3   text ← strip(`text`);

4   symbols = `[#,$, +, -, =, http, https)]`;

5   slangs =`[(åon't", åill not "),("isn't", "is not "),("can't", "can not "),(n't", not "),("i'm", "i am "),("'re", äre "),("'d", åould "),("'ll", åill ")]`

6   **foreach** *word w in text* **do**

7   **if** *contains(word, symbol)* **then**

8   text ← remove_symbol(word, text);

9   **end if**

10   **if** *contains(word, slangs)* **then**

11   text ← replace_slangs(word, slangs);

12   **end if**

13 **end foreach**

14 **return** text

ID	PHQ-9 symptoms	Extracted Lexicons
S1	Little interest or pleasure in doing things	[interest, interested]
S2	Feeling down, depressed, or hopeless	[feeling, depressed, hopeless]
S3	Trouble falling or staying asleep, or sleeping too much	[sleep, asleep]
S4	Feeling tired or having little energy	[tired, energy]
S5	Poor appetite or overeating	[appetite, overeating]
S6	Feeling bad about yourself or that you are a failure or have let yourself or your family down	[failure, family]
S7	Trouble concentrating on things, such as reading the newspaper or watching television	[concentration, reading, watching]
S8	Moving or speaking so slowly that other people could have noticed. Or the opposite being so figety or restless that you have been moving around a lot more than usual	[moving, speaking, restless]
S9	Thoughts that you would be better off dead, or of hurting yourself	[dead, hurt, suicide]

Tabell 1: PHQ-9 symptoms (original) and the extracted lexicons (created)

Algorithm 2: Algorithm to generate lexicons

input : *extracted_lexicons* from PHQ-9

output : Domain specific contextual-aware lexicons

1 seed_terms ← [];

2 **foreach** *symptom s ∈ extracted_lexicons* **do**

3   terms ← [];

4   **foreach** *word w ∈ s* **do**

5   synonyms ← wordnet.synonyms(s);

6   **foreach** *synonym w ∈ synonyms* **do**

7   terms ← wordnet.hyperonym(s);

8   terms ← wordnet.hyponym(s);

9   terms ← wordnet.antonyms(s);

10   **end foreach**

11   **end foreach**

12   sWord ← [];

13   **if** *Model==Depression2Vec* **then**

14   **foreach** *term t in ∈ terms* **do**

15   sWord ← $word2vec_{dep}(t, nWord = 5)$

16   **end foreach**

17   **end if**

18   **if** *Model==Glove2vec* **then**

19   **foreach** *term t in ∈ terms* **do**

20   sWord ← $glove2vec(t, word_{sim} > 80\%)$

21   **end foreach**

22   **end if**

23   **if** *Model==Wordnet* **then**

24   **foreach** *term t in ∈ terms* **do**

25   sWord ← *t*

26   **end foreach**

27   **end if**

28   terms ← sWord;

29   seed_terms ← terms;

30 **end foreach**

31 **return** seed_terms