Fundamentals of Natural Language Project

2022-23

Task: Detection of negation and uncertainty

nº historia clinica: ** *** *** nºepisodi: ********
sexe: dona data de ...

d'hospitalitzacio motiu d'ingres trabajo de parto
antecedents no alergia medicamentosa conocidas ap:
epilepsia en tratamiento no intervenciones quirurgicas
no transfusiones no habitos toxicos medicacio habitual

...

serologias: rubeola inmune, toxoplasma no immune, lues
vih, vhb y vhc negativos. - o'sullivan: 81 - urocultivo:
negativo - cultivos r / v:

...

el 2.08.18 se indica cesarea por sospecha de perdida de
bienestar fetal. a las 20.25 h se obtiene recien nacido
vivo mujer de 3.380 gr, apgar 9(10, ph 7.22-7.27.
hemostasia correcta. sondaje vesical: orina clara.
procedimiento sin incidencias. intradermica en piel. el
pueperio clinico ...

Negation cues

Negation scope

Uncertainty cues

Uncertainty scope

Dataset and annotation


```
{"data":{"cmbd": "null",
         "id": "19062854",
         "docid": "null",
         "page": "null",
         "paragraph": "null",
         "text":" nº historia clinica:..."},
         "annotations":[],
         "predictions":[{"result":
        [{"value":{"start": 347,
                   "end": 350,
                   "labels": ["NEG"]},
                  "id":"ent0",
                  "from_name":"label",
                  "to_name":"text",
                  "type":"labels"},
        {"value":{"start": 350,
                  "end": 372,
                  "labels": ["NSCO"]},
                   "id":"ent1",
                   "from_name":"label",
                   "to_name":"text",
                   "type":"labels"},
        ...
```


nº historia clinica: ** *** *** nºepisodi: ******** sexe: dona data de naixement: 20.06.1999 edat: 19 anys procedencia domicil/res.soc servei obstetricia data d'ingres 02.08.2018 data d'alta 06.08.2018 11:28:06 ates per **********, *****; teixido troyano, anna informe d'alta d'hospitalitzacio motiu d'ingres trabajo de parto antecedents no alergia medicamentosa conocidas ap ...

Dataset and annotation


```
...
{"value":{"start": 2149,
          "end": 2161,
          "labels": ["UNC"]},
                    "id":"ent18",
                    "from_name":"label",
                    "to_name":"text",
                    "type":"labels"},
{"value":{"start": 2161,
          "end": 2188,
          "labels": ["USCO"]},
                    "id":"ent19",
                    "from_name":"label",
                    "to_name":"text",
                    "type":"labels"}
          ...
```


```
...
el 2.08.18 se indica
cesarea por sospecha de
perdida de bienestar fetal
...
```

Detection of negation and uncertainty

Usually, it is approached as a two-step process:

1. Detection of negation/uncertainty cues
2. Detection of the scope of the negation/uncertainty

Methods can be categorized into 2 basic types:

1. Rule-based methods
2. Machine-learning methods

applied sciences

MDPI

Review

Negation and Speculation in NLP: A Survey, Corpora, Methods, and Applications

Ahmed Mahany [1,*], Heba Khaled [1], Nouh Sabri Elmitwally [2,3], Naif Aljohani [4] and Said Ghoniemy [1]

Rule-based methods

Rule-based methods use a set of pre-defined rules to find negation cues and their scope taking into account:

- Pre-defined list of negation trigger words
- Regular expressions
- Part of Speech tagging
- Syntactic parsing

Basic Algorithm: NegEx

- Chapman et al. *A Simple Algorithm for Identifying Negated Findings and Diseases in Discharge Summaries*. Journal of Biomedical Informatics. 2001.
  https://github.com/chapmanbe/negex/tree/master/negex.python

Adaptation to Spanish

- Solarte Pabón et al. *Integrating Speculation Detection and Deep Learning to Extract Lung Cancer Diagnosis from Clinical Notes*. Applied Sciences. 2021
  - Only section on negation and speculation detection
  - Detailed description of the rules created for negation detection in Spanish
- Costumero et al. *An Approach to Detect Negation on Medical Documents in Spanish.*

Machine learning methods

Machine learning methods train classifiers based on text features (PoS, lemma, syntactic features, word embeddings, …) for each of the two sub-tasks (detection of negation/uncertainty signals and detection of the negation/uncertainty scope)

Basic Algorithms

- Enger et al. *A Simple Algorithm for Identifying Negated Findings and Diseases in Discharge Summaries*. 2017.
  https://github.com/marenger/negtool
- Morante et al. *A metalearning approach to processing the scope of negation*. 2009.

Adaptation to Spanish

- Beltrán et al. *Detection of Negation Cues in Spanish: The CLiC-Neg System.* 2019
- Loharja et al. *Negation Cues Detection Using CRF on Spanish Product Review Texts.* 2018

Using Deep Learning (LSTM)

- Fancellu et al. *Neural Networks For Negation Scope Detection.* 2016
- Fabregat et al. *Deep Learning approach for Negation Cues Detection in Spanish*. 2018
- Fabregat et al. *Extending a Deep Learning approach for Negation Cues Detection in Spanish*. 2018

Goal of the project

- **Implement two methods** for detection of negation and uncertainty cues and scopes
  - One method from each category (rule-based and machine learning)
  - No need to follow exactly some of the reference methods. You can create your “own method” based on these references
- **Evaluate** these methods on the provided dataset
  - Standard evaluation metrics: precision, recall, F-score
  - Standard evaluation practice: training/validation/test splits
  - Try to get the best possible results. However, achieving the best performance is not the final goal of the project.

Organization of the project

You will work in **groups of 3/4 students**. You can **register** your group in **Campus Virtual**.

Calendar of the project:

- *April 12th*: Introduction of the project
- *April 26th*: First follow-up of the project
  - Analysis and understanding of the task, dataset and annotation
  - Design of the methods to be implemented
- *May 15th*: Second follow-up of the project
  - Finished implementation and evaluation of rule-based method
  - Started implementation of machine-learning method
- *May 31th*: Final presentation of the project
  - Finished implementation and evaluation of both methods
  - Oral presentation (10 minutes per group) with the main results and conclusions of the project

Organization of the project

Next steps

- Get familiar with the task, the dataset and the annotation format
  - Extract annotations from the JSON file
  - Analyze the annotations: negation and uncertainty cues, rules to determine the scope, …
  - Understand the main challenges of the task
- Get familiar with the reference methods
  - Read the survey paper
  - Read reference papers on rule-based methods: NegEx algorithm and its adaptation to Spanish
  - Start reading reference papers on machine learning methods (non deep learning)
- Design the methods to be implemented
  - Have a look to the links with implementations of the basic methods
  - Think how to adapt the reference methods (specially rule-based methods) to our specific dataset

Questions?

Contact

Ernest Valveny, Ernest.Valveny@uab.cat

Shuxian Pan, pshuxian@iiia.csic.es