ABBREVIATION DISAMBIGUATION WITH NLP

Table of Contents

1. Problem Statement

What is the possibility for a Data Scientist to create a Deep Learning model that can read Medical Text and correctly classify the Medical Abbreviations to enable readers to make sense of the Medical Text in correct context?

2. Context of the Problem

Acronyms and abbreviations within clinical text are widespread, and their use continues to increase. Several reasons for this ongoing growth include adoption of electronic health record (EHR) systems with increased volume of electronic clinical notes accompanied by the wide usage of acronyms and abbreviations, the time-constrained nature of clinical medicine encouraging the use of shortened word forms, and a longstanding tradition of commonly using acronyms and abbreviations in clinical documentation.

The process of understanding the precise meaning of a given acronym or abbreviation in texts is one of several key functions of automated medical natural language processing (NLP) systems and is a special case of word sense disambiguation (WSD).

3. Criteria for Success

The proposed solution should be able to correctly identify the Abbreviation based on the context of the Medical Text.

4. Scope of Solution Space

The proposed solution shall only be applicable to the study of Medical Texts and shall not consider any other aspects of text like causality, Adverse Drug Reactions, etc.

5. Constraints with solution space

- Public availability of precise and Accurate Medical Texts.

6. Data Sources

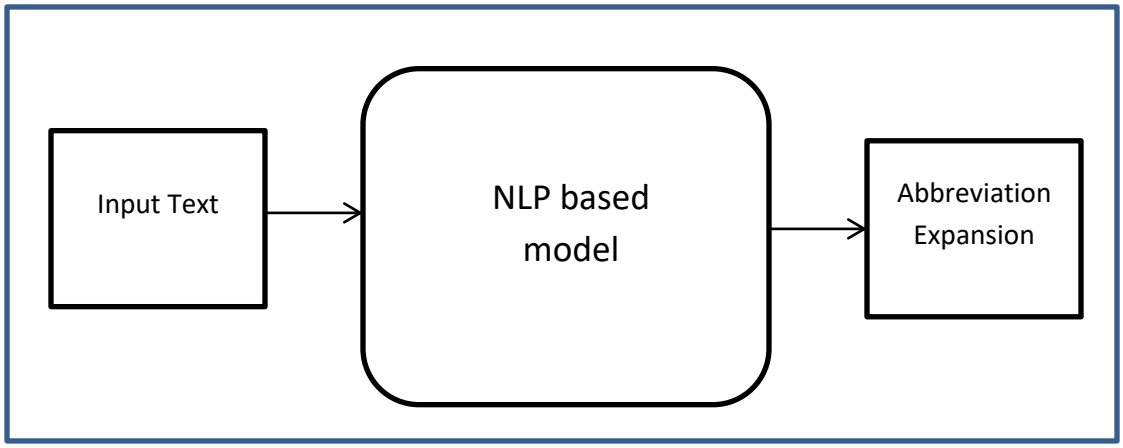
- https://www.kaggle.com/xhlulu/medal-emnlp

7. Stakeholders

- Doctors, Interns, Surgical residents.
- Medical Scholars.

8. Details

- Solution shall consist of a trained deep learning model created by using Natural Language Processing techniques.
- The Input Dataset will consist of Medical Texts with a column containing indexes of the Acronym used in that specific text.
- Since we will be using the Medical Texts and NLP techniques, we won't have to deal with Missing data.



Basic Flow of the Model

9. Deliverables

- Trained Classification Model with code in GitHub Repository.
- A presentation slide deck.
- A Project Report.

10. Conclusion

The proposed solution will create model by using Natural Language Processing techniques which will be used to predict the correct expansion of an Acronym based on the Medical text's context. As a future enhancement, the model can be trained to detect the Drugs and Adverse Reactions in the Medical Journals as well.