

Natural Language Processing

Project 2: Automatic Sense Making and Explanation



Problem Description

Commonsense knowledge and commonsense reasoning play a vital role in all aspects of machine intelligence. In the NLP community, many benchmark datasets and tasks have been created to address commonsense reasoning for language understanding. These tasks are designed to assess machines' ability to acquire and learn commonsense knowledge in order to reason and understand natural language text (Storks et al. 2019).

The objective of this project is to directly test whether a system can differentiate natural language statements that make sense from those that do not make sense. Three subtasks were designed by SemEval 2020.

https://competitions.codalab.org/competitions/21080#learn_the_details-overview

Part 1 (30%)

The first task is to choose from two natural language statements with similar wordings which one makes sense and which one does not make sense;

Example Task A: Validation

Which statement of the two is against common sense?

- Statement 1: He put a turkey into the fridge. (*correct*)
- Statement 2: He put an elephant into the fridge.

Part 2 (30%)

The second task is to find the key reason from three options why a given statement does not make sense;

Example Task B: Explanation (Multi-Choice)

Select the most corresponding reason why this statement is against common sense.

- Statement: He put an elephant into the fridge.
- Reasons:

- **A:** An elephant is much bigger than a fridge. (*correct*)
- **B:** Elephants are usually white while fridges are usually white.
- **C:** An elephant cannot eat a fridge.

Task 1: Sen-Making

Which statement of the two is against common sense?

—	He put a turkey into the fridge	○
—	He put an elephant into the fridge	✗

—	he was sent to a restaurant for treatment	✗
—	he was sent to a hospital for treatment	○

Task 2: Explanation

Why this statement is against common sense?

He put an elephant into the fridge

{

A : an elephant is much bigger than a fridge ✓

B : elephants are usually gray while fridges are usually white ✗

C : an elephant cannot eat a fridge ✗

he was sent to a restaurant for treatment

{

A : a restaurant does not have doctors or medical equipment ✓

B : a restaurant is usually too noisy for a patient ✗

C : there are different types of restaurants in the city ✗

Part 3 (30%)

The third task asks machine to generate the reasons and we use BLEU to evaluate them.

Example Task C: Explanation (Generation)

Generate the reason why this statement is against common sense and we will use BLEU to evaluate it.

- Statement: He put an elephant into the fridge.
- Referential Reasons:
 - i. An elephant is much bigger than a fridge.
 - ii. A fridge is much smaller than an elephant.
 - iii. Most of the fridges aren't large enough to contain an elephant.

Your implementation could be based on the use of language modeling following the work in

Cunxiang Wang, Shuailong Liang , Yue Zhang , Xiaonan Li and Tian Gao. Does It Make Sense? And Why? A Pilot Study for Sense Making and Explanation.

<https://arxiv.org/pdf/1906.00363.pdf>

Evaluation (10%)

Subtask 1 and 2 will be evaluated using **accuracy**. Subtask 3 will be evaluated using BLEU score.

The training data and the evaluation tools for the three sub-tasks can be found on Canvas or via the following link:

<https://github.com/wangcunxiang/SemEval2020-Task4-Commonsense-Validation-and-Explanation>

You need to integrate these tools to your implementation and explain the results of each task.

Submission

You are required to submit:

- An electronic copy of all source code developed (.tar.gz or .zip archive).
- Please put **your student name and number at the top of your file(s)**.
- It is your responsibility to make sure you upload the correct file.

All files can be submitted with Canvas before **22:00 on November 30th 2019**. Please ensure to include **your name and student number** on the **Jupyter notebook document**.

Code

All code should be completed using Python as the programming language.

Your code should have a logical structure and a high level of readability and clarity. Please comment your code and put all code into functions. Your code should be efficient and should avoid duplication.

Late submissions

If you don't get the assignments done to your satisfaction and don't meet the minimum requirements by the deadline, you have the option (as with any assignment at CIT) of submitting up to 1 week late for a penalty of 10%.

This penalty is subtractive. Work that would have earned 55% if on time, would get 45% (not 49.5%) if late.

The penalty is applied weekly. So, 1 day late costs the same 6. If you want to take that option please let me know. Otherwise, I will just correct whatever I have.

If you have a specific reason for submitting a late assignment (sickness, etc) please contact me directly or submit a medical certificate in the department secretary.

Plagiarism

Please read and strictly adhere to the [CIT Honesty, Plagiarism and Infringements Policy Related to Examinations and Assessments](#). Note that reports are **checked** against each other and against external web sources for plagiarism. Any suspected plagiarism will be treated seriously and may result in penalties and a zero grade.

Grading

The assignment is worth 35% of the overall mark for the module. Marks will be awarded based on the quality of the code and the results. In particular, I will be checking to see if you are handling and preprocessing data correctly, carrying out exploratory analysis to gain insights, correctly performing model implementation, and critically, documenting everything in a clear and concise way. The submitted code will also be checked to ensure that the work is your own.