Natural Language Processing Project2 - Semantic Relations

2018

NTU CSIE
授課教師: 陳信希 教授
助教: 顏安孜 陳重吉
{d04922005, d05922016}@ntu.edu.tw

Agenda
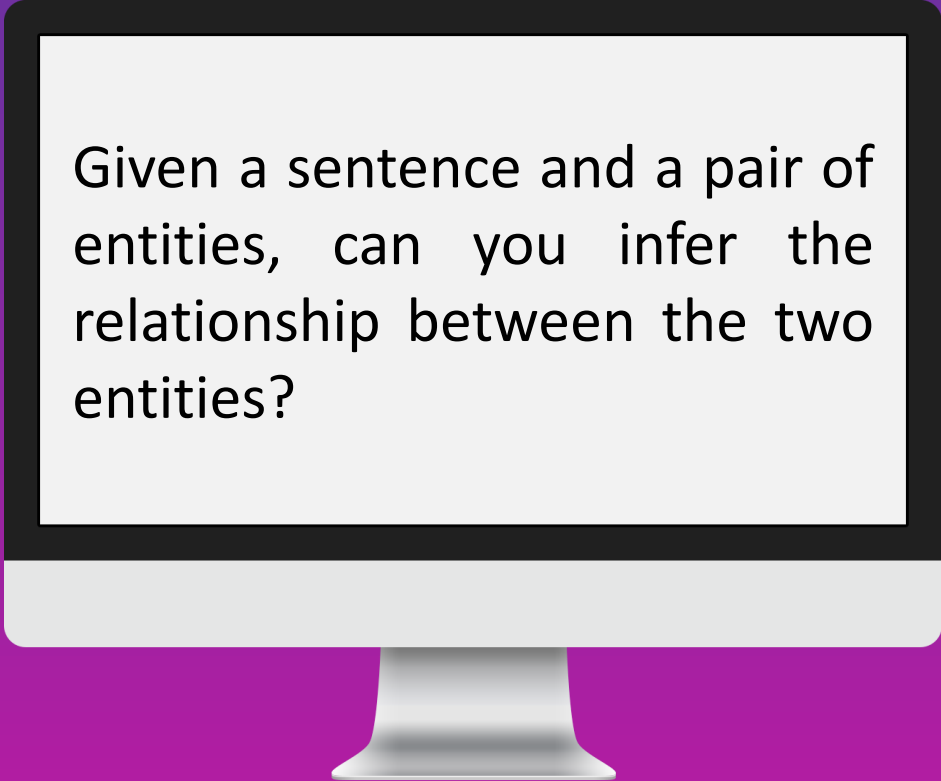
Task Description

Data Format

Evaluation

Report

Important Dates

Task Description

Given a sentence and a pair of entities, can you infer the relationship between the two entities?

SemEval-2010

Task 8:
http://www.aclweb.org/anthology/S10-1006

Nine Relations + Other
Cause-Effect/ Instrument-Agency/ Product-Producer/ Content-Container/ Entity-Origin/ Entity-Destination/ Component-Whole/ Member-Collection/ Message-Topic

Example
We actually did put <e1>**tea**</e1> in the <e2>**food**</e2> Entity-Destination(e1,e2)

Cause-Effect (CE). An event or object leads to an effect.
Example: those cancers were caused by radiation exposures

Instrument-Agency (IA). An agent uses an instrument.
Example: phone operator

Product-Producer (PP). A producer causes a product to exist.
Example: a factory manufactures suits

Content-Container (CC). An object is physically stored in a delineated area of space.
Example: a bottle full of honey was weighed

Entity-Origin (EO). An entity is coming or is derived from an origin (e.g., position or material). Example: letters from foreign countries

Entity-Destination (ED). An entity is moving towards a destination.
Example: the boy went to bed

Component-Whole (CW). An object is a component of a larger whole.
Example: my apartment has a large kitchen

Member-Collection (MC). A member forms a nonfunctional part of a collection.
Example: there are many trees in the forest

Message-Topic (MT). A message, written or spoken, is about a topic.
Example: the lecture was about semantics

Other

Data Format - Train

TRAIN_FILE.txt

ID   Sentence

Relation

Comment

Ex:

3   "The <e1>author</e1> of a keygen uses a <e2>disassembler</e2> to look at the raw assembly code."

Instrument-Agency(e2,e1)

Comment:

5   "The <e1>student</e1> <e2>association</e2> is the voice of the undergraduate student population of the State University of New York at Buffalo."

Member-Collection(e1,e2)

Comment:

Data Format - Test

TEST_FILE.txt

ID   Sentence

Ex:

8001   "The most common <e1>audits</e1> were about <e2>waste</e2> and recycling."

8002   "The <e1>company</e1> fabricates plastic <e2>chairs</e2>."
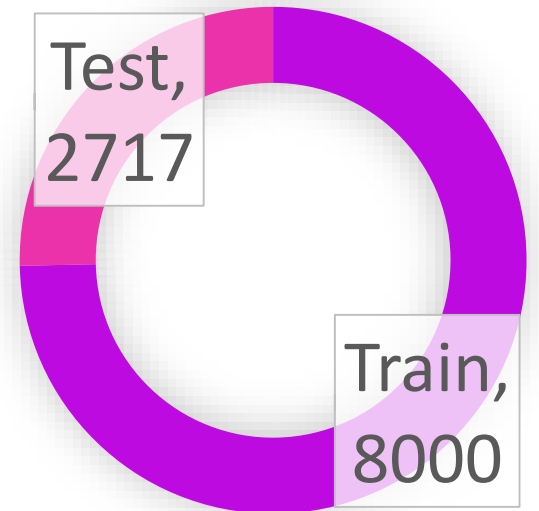
▲

Format of Proposed Answer

<ID>   <RELATION>

with a TAB as a separator
e.g.,
- 1   Component-Whole(e2,e1)
- 2   Other
- 3   Instrument-Agency(e2,e1)

Data

- Train
- Test

Test, 2717
Train, 8000

▲
F1-score

Use the provided scorer:
perl semeval2010_task8_scorer-v1.2.pl <PROPOSED_ANSWERS> answer_key.txt

Examples:
perl semeval2010_task8_scorer-v1.2.pl proposed_answer1.txt answer_key.txt > result_example.txt

*Details of the output file are in the **result_example.txt**

Report

- Language: Chinese or English (Be clear in meaning!)
- Pages: no more than 6 pages (with readable font size)
- Format: PDF
- Must include:
  - Name and student ID of each team member
  - Agree to share your report with your classmates? (YES/NO)
    - Will appear in 作業觀摩
  - Methodology
  - Experiments
  - Discussions

Submission

- Deadline: 2018/06/29 23:59

- Upload your report to CEIBA
  - Provide the link of code (ex: GitHub) in the report
  - One submission per group

- Grading:
  - Performance: 30%
  - Report: 70%
  - Bonus: 10%
    If your performance beat the best result in the paper.

Reference
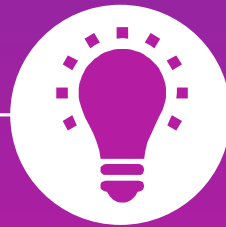
- [SemEval 2010](#)
  - S10-1043~S10-1050

Important Date

2018/05/31
Project 2 Release

2018/06/29 23:59
Submit report to CEIBA

Good
Luck