CPSC 532V: Common sense Reasoning in NLP
Assignment #1

Sadaf Sadeghian

Step 1)

I chose ConceptNet KB for this assignment.

Reading Data & Preprocessing:

I first downloaded the ConceptNet assertions (edges) locally, then read the file line by line and filter it for English words. Using the extracted English dataset, I constructed a list of tuples, each tuple having the information of an edge in the graph and constructed as bellow:
(source, target, {'relation': type_of_relation, 'weight': weight_value})

For Example, a tuple constructed using this method is:

('abandon', 'gain', {'relation': 'Antonym', 'weight': 1.0})

Constructing Graph:

I used the network library for this part. Using the list of tuples constructed in the previous phase, I constructed a directed graph. Then, the reverse edges were added to the graph with the same weight value, and type of relation is "relation -1" (For instance "RelatedTo -1").

Finding Paths:

For finding the path between two given words, I first found the length of the shortest path between the pair of words in the graph. Later, I found all the paths of that length (length of the shortest path between the pair) and choose the best path based on the product value of the weights of the edges in that path. Assuming the path which has the highest product value of weight of edges is the most reliable path.

* *nx.shortest_path_length* and *nx.all_simple_paths* from network library is used in this step.

Visualizing Paths:
I used the graphviz library for visualizing graph representation of each path separately.

Efficiency: Downloaded the KB locally
Node matching: Exact (verbatim) Matching
Reversed edges: Added (same weight value, and type of relation is "relation-1").
Edge scores: Used in finding the best path among the paths with the same length

Step 2)

All of the paths are shown in the jupyter notebook in the Examples part. There is a subsection for each example. (Also, in the path_graphs.pdf)

Step 3)

1) Which types of knowledge are the most common?

I analyzed the occurrences of each type of knowledge (relation in ConceptNet) among the founded paths between all the pairs for the examples. Based on the result the top three most common types of knowledge are “RelatedTo”, “HasContext” and “IsA” with their reversed edges.

In the following I show the number occurrences of each one:

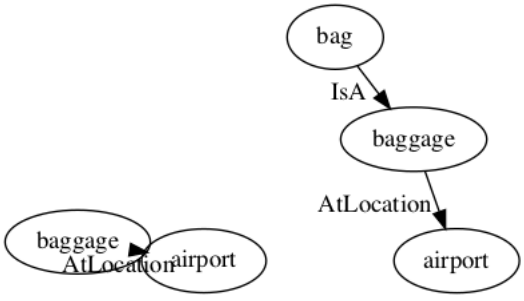
```
'RelatedTo': 441
'RelatedTo -1': 388
'HasContext': 94
'HasContext -1': 91
'IsA': 78
'IsA -1': 76
```

2) Which types of knowledge are the most accurate / reliable?

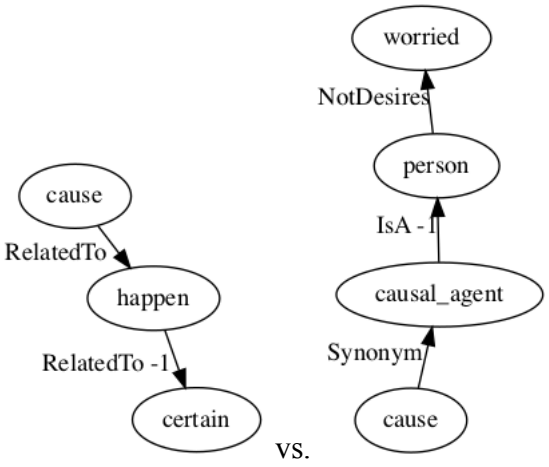
In ConceptNet, for each assertion or edge in the graph, weight attribute indicates the strength of this assertion and how reliable and accurate the edge is. However, based on my observation, the synonym, antonym and AtLocation knowledge types were the most accurate because they are more obvious and cannot interpret differently. In the other hand, HasContext type of knowledge was the one I thought it’s mostly misleading, some misleading examples:
Has context (us, garbage) (us, tie) (medicine, host)

3) Which types of knowledge are the most useful for each instance (i.e. contain information useful for answering the question)?

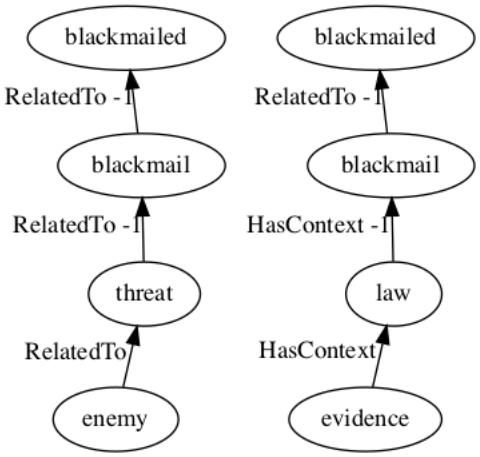
Example 1) IsAt (Because we are asking about a location which is the place she is heading)



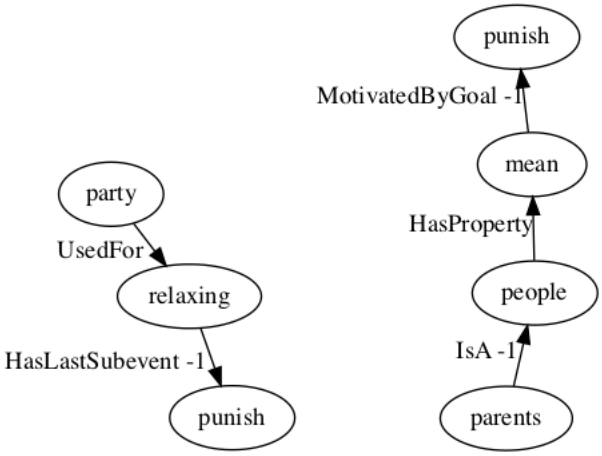
Example 2) RelatedTo (Searching for the cause of event and want to choose between being certain or worried about catching flu. We can see the correct answer has a shorter path)



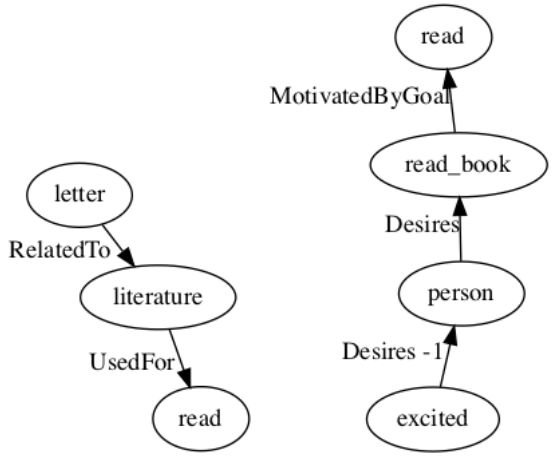
Example 3) RelatedTo and HasContext **



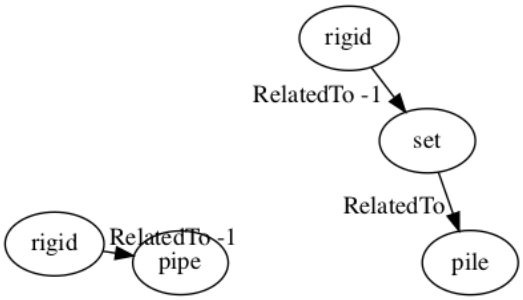
Example 4) The paths for this example is not that good but based on the two important paths shown below HasProperty and MotivatedByGoal (Because we are talking about a social behavior of people)



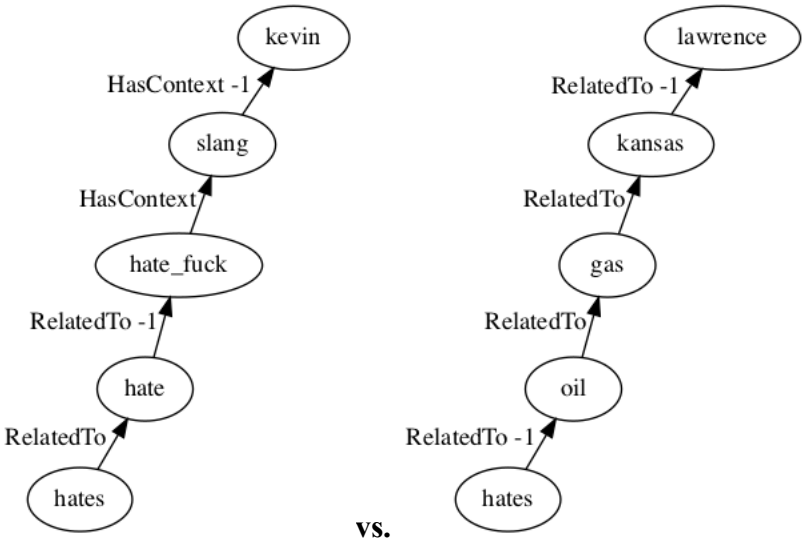
Example 5) Desires and MotivatedByGoal (Because we are talking about a social behavior of people)



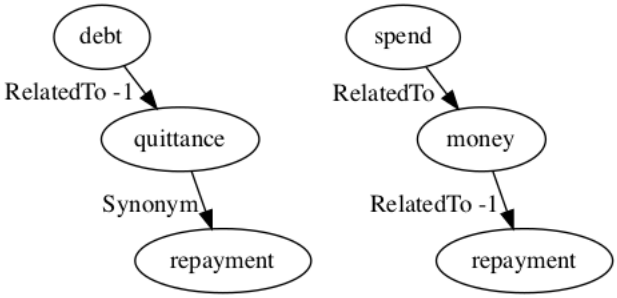
Example 6) RelatedTo (We are talking about which one is rigid and which one is more related to rigid, and we can see a shorter path between rigid and pipe rather that rigid and pipe)



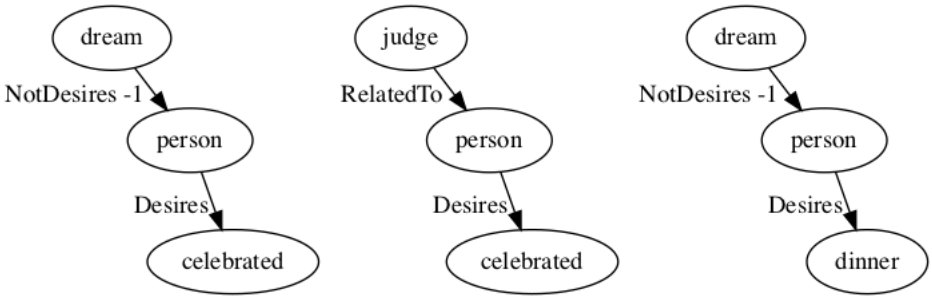
Example 7) no relation is helpful because interpreting the order of words and parts of speech in the sentence is more important in this example than paths and relationship between pairs of words because the choices are just two names…. As you can see below the paths are not helpful at all



Example 8) RelatedTo however since there is a cause and effect relationship between the sentence I expected the Causes knowledge type but none of the paths are using this type of knowledge for this example for now



Example 9) This approach is not working that well for this example but the most useful paths are these three paths and based on them Desires



Example 10) This approach is not working that well for this example since it is describing a process and sequence of actions which is not reflected in our method which only considers word pairs

4) Do you recognize any problems with the knowledge or path search approach? What can be improved?

Yes, it only considers pairs of words which can be helpful in cases where the words of the answer are so related to the question and other choices are not that related (for instance Example1) However, this method cannot do inference based on the order of actions, cause-effect relationships, grammar structure of the sentence, …. Which makes it not helpful in more advance Q&A tasks which require more complex inference where the answer is not interpretable by pair of words (for instance PIQA, WinoGrande, …).