

Review #1

What is this paper about, what contributions does it make, what are the main strengths and weaknesses?

This paper promises to compare the effort needed to extract classical encyclopedic knowledge compared with commonsense knowledge (i.e., the former being about instances, the former about classes of concepts). As such the topic is of potential relevance to ACL. CSK is a topic of rising importance, and a careful analysis into extraction challenges could be a useful contribution.

My main issue with the paper is that it does not stand up to its promise. Instead of encyclopedic knowledge and commonsense knowledge, it compares just two isolated resources, Freebase and ConceptNet, on general text corpora. As such, it is highly prone to overfit to the idiosyncrasies of the two resources (which are also somewhat dated), and the input text corpora.

At the same time, the paper ignores import advances on KB construction, in particular Wikidata [1], TupleKB [2] and Quasimodo [3]. If the paper would take efforts to show that the two resources are representative for their classes this limitation might be acceptable, however, the paper takes no such efforts, and especially for ConceptNet this reviewer is quite certain that this is not the case. ConceptNet is a mashup of >90% automatically imported lexico-syntactic data, while less than 10% of it relate to what is typically considered commonsense. Similarly, it uses only ~20 predicates which are arguably rather meta-predicates (the objects of relations like "can be used for", or "has precondition" typically themselves contain other predicate phrases), so at this meta-predicate sparsity, it is no wonder that meta-predicates are overloaded. TupleKB [2] and Quasimodo [3] here provide informative counterpoints, relying on OpenIE instead of fixed meta-predicates, and thus having thousands and more predicates, which should lead to completely different characteristics.

By using general text corpora the paper also ignores the established insight that an important step in CSKB construction is careful source selection (as emphasized by all of WebChild, TupleKB, Quasimodo).

As it stands, I do not see a message that extends beyond two idiosyncratic resources, and thus recommend rejection.

Reasons to accept

Brings an interesting topic to discussion - what distinguishes commonsense from encyclopedic knowledge.

Reasons to reject

Indication that lack of thoroughness and overselling is acceptable at ACL.

Overall Recommendation: 1.5

Missing References

- [1] Vrandečić, Denny, and Markus Krötzsch. "Wikidata: a free collaborative knowledge base." (2014).
- [2] Mishra et al. "Domain-targeted, high precision knowledge extraction." Transactions of the Association for Computational Linguistics 5 (2017): 233-246.
- [3] Romero et al. "Commonsense Properties from Query Logs and Question Answering Forums." CIKM (2019).
- [4] Chaliar et al.. "Joint Reasoning for Multi-Faceted Commonsense Knowledge." arXiv preprint arXiv:2001.04170 (2020).

Typos, Grammar, and Style

Minor comments

- The title is misleading, the paper does not investigate a binary question
- Please load English babel package, then odd breaks like "s-tudy" (line 15) should go away
- Paper requires careful proofreading, e.g. "commonsensesense"
- L102: What about OpenIE?
- L130: What is "side information"?
- L142: Which subset? What is the selection rationale?
- L203-212: The argument that CSK has broader and more ambiguous semantics is sensible. Yet I do not see this taken up later - a sensible way to study it would be, e.g., annotator agreement. Dice [4] also has an interesting analysis on the semantics of CSK.
- "can't" is uncommon in professional writing
- L322: Argument unclear
- L323-L332: This subdivision was an insight of the WebChild project, unlike you really came up with that yourself, please explain that this is not your contribution. Also the argument is quite unclear, Wikidata has over 5000 relations, and nonetheless, its "part of" faces the same issue. Why is it actually an issue?
- Line 345ff: The metrics in table 7 do not support the textual description. Unless for DT, there is hardly any difference between F and C, and even for DT, a 50% increase does not support that CSK is so much harder.

Review #2

What is this paper about, what contributions does it make, what are the main strengths and weaknesses?

This paper presents a linguistic study of commonsense relations in natural language text.

The authors distinguish commonsense knowledge from factual knowledge in the introduction. Here, I wondered whether "commonsense knowledge" could also be called probabilistic knowledge or generalised knowledge? In which application area are probabilistic relations particularly useful (they certainly are in various reasoning tasks, but a more detailed explanation of why this work is relevant would have been nice, and would certainly have added value to the paper)?

In section 3, a short explanation of OpenNRE is in order. I, for instance, did not understand whether relation discovery or only relation classification is performed. Examples of the 54 and 27 relation types would have been nice. Are they the same across the different corpora? An explanation of MRR is missing.

The authors argue that factual RE is much harder than commonsense RE. However, Wiki (FB) performance is far from being good. I think the authors fail to prove their claim, namely that the alleged properties of commonsense relations really affect system performance.

I dare say that the authors' findings are not very surprising.

Reasons to accept

It's a study on a relevant topic, but in my view, this paper should not be accepted.

Reasons to reject

The authors fail to prove their claim.

Overall Recommendation: 2

Typos, Grammar, and Style

I think the document language is not English. There are numerous incorrect line splits. A language review is needed.

"removal of NA relation" in lines 163 ff. is unclear.

Layout on page 3 is not good (leftover line from section 3).

Review #3

What is this paper about, what contributions does it make, what are the main strengths and weaknesses?

The paper discusses the assumption that extracting common sense from texts is harder than extracting facts. The main weakness of the paper is its main goal has been quite obvious. Nobody would expect something different from that. By their nature, concepts may be described using different NL phrases. Facts about specific entities are easy to extract from text precisely because we are talking about names that tend to have fewer variations. But the most important point is that no real information is added at this level of generality. One could have to dig into particular classes of concepts and entities to better classify the complexity of the extraction of references to it.

Reasons to accept

The paper presents some datasets that could be presented for the community. But no important result or conclusion.

Reasons to reject

The conclusions are obvious. It is not clear what one can learn from the paper.

Overall Recommendation: 2.5

Questions for the Authors(s)

Obama is Canadian. This is a fact with one argument being a NE. Is it absolutely true? Please make clear the introduction. I didn't understand your notion of "absolute true" or "always false" for facts.

Please consider refine your statement about YOUR discoveries about concept polymorphism, concept polysemy, and relation polysemy. This is a well-known fact since books like Natural Language Understanding (2nd Edition, 1994) from James Allen. Moreover, your experiments are using very specific samples of texts: NYT (facts), Wikipedia (common sense), and literature (more neutral but mostly facts about fictional people, but more have to be said about the books you selected).

Section 3. you wrote, "first we use all the triples in CN* to retrieve input sentences from the RC model". Can you explain how? Later, you talk about head and tail of triples... Please define. Normally head and tail are defined for lists and triples are not lists. Moreover, negative examples could be defined not only for random entities in the KB but also for nonexistent entities. That is, what is your notion of *negative example*?

The conclusions from Table 2 you say that numbers support the conclusion that extracting common sense is much harder than facts. But you haven't explained the model part. you just said that used GloVe but how? Please cite and explain the PCNN+ATT, the intuition behind it is very related to your conclusions.

Section 4.1. Please define entropy in the context of your work. As said above, it is quite obvious that concepts are more polymorphic than names.

Section 4.2. "When the number of words of a concept increases, the number of possible senses decreases." This is once more quite well-known. A complex noun-phrase is a specialization of a concept: "a dog" vs "a black dog", "a black dog holding a bone in its teeth" etc. What is the insight here? Maybe you need to be more clear about what is the discoveries that you made.

Table 7: please can you define DPL and DT? Can you say what tool you used for parse and POS tagging (section 4.2 mentioned POS tags).

Missing References

citation to the tools used for parsing and POS tagging.

Typos, Grammar, and Style

Introduction. Consider avoiding the hyphenation of `tex-t`

Additional Suggestions for the Author(s)

you mentioned the word2vec from Lin et al. But ended up saying that you used GloVe embedding from Pennington et al. So why the reference to the first? Why not go directly to the description of the second?

Please explain the MRR rank.

In the example `baking`, it can be both the gerund of the verb `bake`, its nominalization (the process). Or the `present participle` form of the verb. Not sure after this example how can I take your numbers and conclusions. Actually, it was not clear what is your conclusion about the single/multiple words in the triples. Section 4.2 is not clear.

The assertion "this shows that common-sense relations inhabit in much more diverse linguistic context" is driven by the numbers from Table 7. I claim this assertion is very strong and should be avoided. All your results came from particular corpora and we could think on many ways to construct elaborated syntactical structures for facts. Consider, for instance, domain-specific corpus, from deeply technical domains such as geology.

Review #4

What is this paper about, what contributions does it make, what are the main strengths and weaknesses?

This paper presents a preliminary study on identifying reasons why typical relation extraction models that work reasonably well on factual knowledge extraction fail on commonsense relations.

The paper suggests several possible reasons and give some empirical evidences on why techniques used to extract factual relations fail to extract common sense relations

Drawing a sample from different datasets, the study investigates the problem associated with extracting common sense relations. The analysis is carried out on a small sample drawn from the relation dataset.

The study employs an on-the-shelf tool that performs well on commonsense relation extraction task.

The study broadly analyse three different areas which impacts on extracting commonsense relations as against factual relations.

Reasons to accept

The paper presents a good study on examining the reasons why relation extraction is difficult on common sense relations.

At the same time, the paper distinguishes on different reasons as to why relation extraction that works quite well on factual knowledge fails to work on commonsense knowledge.

The objective of the study is well defined, and the methodology is clearly formulated.

Reasons to reject

The main problem with the paper is that whether the contribution made in the paper is enough to be accepted to ACL.

The paper states that it is a preliminary study and I suspect whether the small sample size drawn for the analysis is good enough to draw broader conclusions about the commonsense relations and the dataset in general.

The paper also does not address on how to solve the problems seen in extracting commonsense relations.

Overall Recommendation: 3.5