Modelling Generics

SAARLAND UNIVERSITY

Name: Priyanka Upadhyay (Master’s DSAI student)

Matriculation: 2581714

Email ID: s8prupad@stud.uni-saarland.de

Seminar: Commonsense Knowledge Extraction and Curation

Agenda

GenericsKB: A Knowledge Base of Generic Statements

Sumithra Bhakthavatsalam, Chloe Anastasiades, Peter Clark [1]

Contents

Background

➢ **What is a Generic Statement?**

- Statements which expresses regularities and lead to facts

E.g.-"Potatoes contain vitamin c, amino acids, protein and thiamine".

- Do not report about specific but shows a general property (not unique).

E.g.- "A duck lays eggs".

But it does not show how many members of category have the property.

E.g. – "Ravens are black" (Carlson 1977).

Problem

- The current AI lacks many features of human commonsense reasoning. This means AI often makes mistakes than human.

E.g.- your domestic robot is at home and you are late from office. Robot has to feed the kids and there is nothing in the fridge. Robot seems the cat and the robot has not learned the human value function properly so it does not understand the sentimental value of the cat outweighs the nutritional value of the cat and then “Deranged robot cooks Kitty for family dinner” and the one incident would be the end of the domestic robot industry - [Stuart Rossel]

E.g.- Existing self-driving cars cannot reason about the location nor the intentions of pedestrians in the exact way that humans do, and instead must use non-human modes of reasoning to avoid accidents

- In NLP, it has been a constant problem to search for additional repositories of general/commonsense knowledge to boost performance further (see prior work in next slide).

Contents

Introduction

Research Goal:

- A good approach: Create a GENERICSKB which can be used in NLP filed.
- Prior works: The most used repository ConceptNet and WordNet are limited but GENERICSKB is the first repository which contains natural generic sentences which are semantically complete sentences.
- Authors have also created GENERICSKB-BEST.

Contents

Dataset

GENERICSKB: GENERICSKB contains over 3.4M sentences which are taken from three corpora:

I. The Waterloo corpus.
II. SimpleWikipedia.
III. The ARC corpus.

BERT Classifier functioning:

- Dataset split into 70:10:20 as train:dev:test instances.
- Resulted statements are 3.4M generic semantic sentences.

GENERICSKB-BEST:

- It contains the best-quality generics from GENERICSKB and finest, synthesized generics from WordNet and ConceptNet.

Contents

Approach

To create GENERICKB 3 steps would be followed:

I. **Cleaning:** Clean the source data.
II. **Filtering:** Find likely generic sentences.
III. **Scoring:** Find semantic generic sentences by applying a BERT classifier.

Corpus	Size (# sentences) Original	Cleaned	Filtered
Waterloo	~ 1.7B	~ 500M	~ 3.1M
SimpleWiki	~ 900k	~ 790k	~ 13k
ARC	~ 14M	~ 6.2M	~ 338k
GENERICSKB	~ 1.7B	~ 513M	~ 3.4M

Figure 2: Corpus sizes at different steps of processing [1].

1. Cleaning:

- Regular Expressions.
- Sentence, token length heuristics to separate malformed sentences.
- Text cleanup using the Fixes Text For You.

2. Filtering (use 27 rule to Identify Generics):

- Sentences start with a bare plural – E.g., “Dogs are...” are considered as good generics.
- Sentence start with a determiner – E.g., “A man said...” are not considered as good generics.

no-bad-first-word: Sentence does not start with a determiner ("a", "the",...) or selected other words.
remove-non-verb-roots: Remove if root is a non-verb
remove-present-participle-roots: Do not consider any present participle roots.
has-no-modals: Sentences containing modals ("could", "would", etc) are rejected
all-propn-exist-in-wordnet: All (normalized, non-stop) words are in WordNet's vocabulary

Figure 3: Sample of filtering rules [1].

3. Scoring:

- BERT classifier used to find semantics generics.
- 3 options were given to crowd workers to provide their suggestions about a statement – yes, no, unsure.

E.g.- Statements about people and firms. Like - Apple makes plenty of iPhones.

BERT Classifier functioning:

- Each fact was scores (yes/unsure/no = 1/0.5/0).
- Dataset split in 70:10:20 into train:dev:test, and a BERT classifier fine-tuned on the training set.
- Outputs representing the two classes (yes/no), followed by a softmax to get class probabilities
- This classifier scored 83% on the held-out test set.

GENERICSKB's Example

1. Example generics about "tree" in GENERICSKB

Trees are perennial plants that have long woody trunks.

Trees are woody plants which continue growing until they die.

Most trees add one new ring for each year of growth.

Trees produce oxygen by absorbing carbon dioxide from the air.

Trees are large, generally single-stemmed, woody plants.

Trees live in cavities or hollows.

Trees grow using photosynthesis, absorbing carbon dioxide and releasing oxygen.

2. An example entry, including metadata

Term: tree

Sent: Most trees add one new ring for each year of growth.

Quantifier: Most

Score: 0.35

Before: ...Notice how the extractor holds the core as it is removed from inside the hollow center of the bit. Tree cores are extracted with an increment borer.

After: The width of each annual ring may be a reflection of forest stand dynamics. Dendrochronology, the study of annual growth rings, has become prominent in ecology...

Figure 1: Example of generic statements in GENERICSKB plus one showing associated metadata [1]

GENERICSKB-BEST:

- GENERICSKB-BEST contains 1,020,868 generics (774,621 from GenericsKB plus 246,247 synthesized from ConceptNet and Wordnet).
- It contains the best quality generics in it’s repository.

Contents

Evaluation

1. Question-Answering:

Corpus	Size	Score on OBQA (test)
QASC-17M	17M	0.660
GENERICSKB	3.4M	0.632
GENERICSKB-BEST	1M	**0.678**

Figure 4: Comparative performance of different corpora for answering OBQA questions [1].

The results indicate that using the high-quality version of GenericsKB-Best it can lead to improved QA performance over using the first corpus.

2. Explanation Quality:

Example: What can cause a forest fire?

storms because:Storms can produce lightning

AND Lightning can start fires

Corpus	Explanation Quality on OBQA	Explanation Quality on QASC
QASC-17M	0.44	0.66
GENERICSKB-BEST	0.61	0.79

Figure 5: Comparative quality of two-hop explanations (sentence chains), generated using two different corpora for two different question sets [1].

Contents

1. Introduction : Research Goal and prior work
2. Dataset
3. Approach
4. Evaluation
5. **Summary**

Summary

- Authors tries to address importance of generics in NLP and tries to provide theory of Genericity.
- Motivation: method to create a repository which contain Generic sentences and it’s use can boost performance in NLP research area - GENERICSKB
- Create GENERICSKB-BEST which contains generics collected from ConceptNet, Wordnet and GENERICSKB.
- Evaluation: compares and claim to be that GENERICKB contain high value generics which can be used in Question-Answering and Explanation Quality.

Thank you

Discussion

- The authors could have tried to provide more clearer results from filtering because after probing all the processes in my perspective filtering rules are a touch ambiguous, it could be more clear to produce better ends up in terms of providing semantic generics.
- Authors could have also given more results of Question-Answering evaluation with few examples to clear reader viewpoint.
- Authors could have shared their approach to create on GENERICSKB-BEST.
- Authors could have included their detailed work to see the original dataset accuracy and to understand the work in better way.

References

[1] Sumithra Bhakthavatsalam, Chloe Anastasiades, and P. Clark. Genericskb: A knowledge base of generic statements. ArXiv, abs/2005.00660, 2020.

[2] G. Carlson. On the semantic composition of english generic sentences, 1989.

[3] Andrei Cimpian and E. Markman. The generic/nongeneric distinction influences how children interpret new information about social others. Child development, 82 2:471–92, 2011.

[4] Peter Clark, Isaac Cowhey, Oren Etzioni, Tushar Khot, A. Sabharwal, Carissa Schoenick, and Oyvind Tafjord. Think you have solved question answering? try arc, the ai2 reasoning challenge. ArXiv, abs/1803.05457, 2018.

[5] Peter Clark, Oren Etzioni, Tushar Khot, A. Sabharwal, Oyvind Tafjord, Peter D. Turney, and Daniel Khashabi. Combining retrieval, statistics, and inference to answer elementary science questions. In AAAI, 2016.

[6] B. Hale and C. Wright. A Companion to the Philosophy of Language, chapter 18. 1997.

[7] Tushar Khot, Peter Clark, Michal Guerquin, P. Jansen, and A. Sabharwal. Qasc: A dataset for question answering via sentence composition. In AAAI, 2020.

[8] Karen T. Kohl, D. Jones, and R. Berwick. Wordnet. an electronic lexical database. edited by christiane fellbaum. 2003.

[9] MIT Media Lab. ConceptNet an open, multilingual knowledge graph, 1999.

[10] Todor Mihaylov, Peter Clark, Tushar Khot, and A. Sabharwal. Can a suit of armor conduct electricity? a new dataset for open book question answering. In EMNLP, 2018.

[11] George A. Miller. WordNet a lexical database for english, 1995.

[12] Robyn Speer, J. Chin, and Catherine Havasi. Conceptnet 5.5: An open multilingual graph of general knowledge. ArXiv, abs/1612.03975, 2017.

[13] . Gregory n. carlson and francis jeffry pelletier (eds.), the generic book, chicago: The university of chicago press, 1995., x+463pp.

[14]. 3 steps to create safer AI -https://plato.stanford.edu/entries/generics/