Meaningful relations in a common-sense knowledge graph

Guy Aglionby and Simone Teufel
Department of Computer Science and Technology
guy.aglionby@cl.cam.ac.uk

Introduction

- We introduce a new structure for knowledge graphs that store common-sense information. This comprises a set of relations based on semantic roles that aim to be highly generalisable across multiple tasks and domains.
- Common sense is the information that most people know and use to understand and reason about the world. It is generally true and widely-applicable, but is defeasible (can be overruled in the presence of more specific information). Machine learning systems must have access to this to be able to complete reasoning tasks.
- One way of storing information is in a knowledge graph. This has the benefit of being an explicit representation, rather than uninterpretable latent information stored in a neural network.
- Knowledge graphs store information as *(subject, relation, object)* triples; although this is too simple to model many linguistic phenomena, it allows the use of prior work on graph representation learning.
- The largest-scale existing knowledge graph of this kind is ConceptNet (Speer et al., 2017; Liu and Singh, 2004). We examine cases where the relations used there may suffer from ambiguity and propose an alternative set to alleviate this.
- We annotate a science question answering dataset, WorldTree (Jansen et al., 2018), in this formalism to demonstrate its feasibility. The result is a dataset where each question is paired with a set of triples from a knowledge graph which jointly explain the correct answer.

WorldTree Translation

- WorldTree stores the facts required to answer multiple-choice questions in a set of tables. Each fact is a row, and each question is labelled with a set of rows.
- There are 62 tables, each representing a different relationship type. Some are domain-specific (*habitat, lifespan*) while others are more general (*kind of, affect*).

Qualifier	Organism	Habitat
some	animals	forests
	fish	water

Table 1: Extract from the *habitat* table.

- Each table has a unique set of columns – the number of these quickly grows as more complex facts are stored, and there is difficulty in ensuring the same class of object populates each column.
- The table-based method of storing information is domain-specific and not scalable; translating tables into a single knowledge graph alleviates these issues while preserving the fact-level labelling.
- These labelled edges can then be used as supervision signal for a question answering model. Few existing datasets provide this level of supervision.

Question
Kim wants to break apart some potassium. Which of the below should they use to do this?

Answers
a) Water
b) Knife
c) Saw
d) Bunsen burner

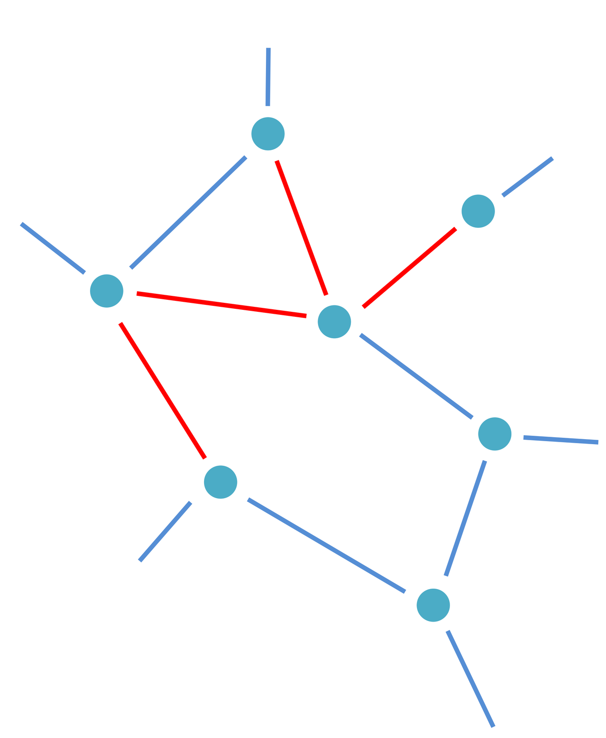

Figure 1: All of the supervision available to a model is shown in red. In addition the correct answer, which is usually the only label available, a set of relevant relations in the graph are also selected.

Knowledge Graph Ontology

- The semantics of each relation should be well-defined and domain-agnostic to maximise generalisability.
- In this work we use semantic roles in the style of FrameNet (Baker et al., 1998) to link verbs and the arguments they can plausibly take in a bipartite graph.
- We contrast this with the relations used by ConceptNet, each of which covers a broader range of relationship semantics.
- Arguments are joined in a taxonomy imported from WordNet (Miller, 1995).
- A subset of the relations we use are shown below (definitions adapted from Jurafsky and Martin (2020)).

Relation	Definition
Agent	The volitional or non-volitional causer of an event.
Patient	The undergoer, experiencer, or entity moved by the event.
Instrument	An entity used in an event.
Result	The end product of an event.
Benefactive	The entity that benefits from the event.
Source	The origin of an entity in a transfer event.
Goal	The destination of an entity in a transfer event.

Table 2: Definitions of some relations in the proposed framework.

- The specificity of the definitions makes precise classification of the relationship between two objects straightforward.
- As shown below, in ConceptNet there is ambiguity between whether a knife is *used for* or is *capable of* cutting and slicing. These two events are of the same type and so should have the same relationship with 'knife'.
- The proposed framework makes the it is clear that a knife is not capable of autonomously cutting and so does not take the agent role; in most cases is only part of a cut or slice event when used as an instrument.
- The ConceptNet node *dust furniture* does not encode that *furniture* is the object of *dust*, and compounds into one node what can be expressed more precisely using two triples with *patient* and *troponym* relations.
- In the last case, the relation *has subevent*, which applies in a broad range of circumstances and here uses another compound node *make food*, is simplified with the more specific relation *result*.

Proposed framework	ConceptNet
(write, *instrument*, pen)	(pen, *used for*, write)
(slice, *instrument*, knife) (cut, *instrument*, knife) (slice, *agent*, person)	(knife, *used for*, slice) (knife, *capable of*, slice)
(walk, *beneficiary*, dog)	(dog, *capable of*, walk on leash)
(dust, *patient*, furniture) (dust, *troponym*, clean)	(dust furniture, *has subevent*, clean)
(cook, *result*, food)	(make food, *has subevent*, cook)

Table 3: Example triples in the proposed framework compared with ConceptNet equivalents.

Future work

- The hypothesis that the reduced ambiguity of the proposed ontology is useful for machine learning models must be tested. We will carry out experiments on the WorldTree dataset with both the annotated relations and a ConceptNet-based equivalent. We will also explore using another dataset for further comparison.
- We will develop a model which is able to use the graph-level supervision now available in WorldTree.
- For the proposed ontology to be applicable to a broader range of problems, its coverage must be expanded. We will explore automated methods of doing this using resources like FrameNet, Open Mind Common Sense (Singh et al., 2002), and OpenCyc (Lenat, 1995).

UNIVERSITY OF CAMBRIDGE

References

Baker, C. F., Fillmore, C. J., & Lowe, J. B. (1998). The Berkeley FrameNet Project. *COLING*.
Jansen, P., Wainwright, E., Marmorstein, S., & Morrison, C. (2018). WorldTree: A Corpus of Explanation Graphs for Elementary Science Questions supporting Multi-hop Inference. *LREC*.
Jurafsky, D., & Martin, J. H. (2020). *Speech and Language Processing* (3rd (draft)).
Lenat, D. B. (1995). CYC: A large-scale investment in knowledge infrastructure. *Communications of the ACM*, *38*(11), 33–38.
Liu, H., & Singh, P. (2004). ConceptNet—A Practical Commonsense Reasoning Tool-Kit. *BT Technology Journal*, *22*(4), 211–226.
Miller, G. A. (1995). WordNet: A Lexical Database for English. *Commun. ACM*, *38*(11), 39–41.
Singh, P., Lin, T., Mueller, E. T., Lim, G., Perkins, T., & Li Zhu, W. (2002). Open Mind Common Sense: Knowledge Acquisition from the General Public. In R. Meersman & Z. Tari (Eds.), *On the Move to Meaningful Internet Systems 2002: CoopIS, DOA, and ODBASE*. Springer Berlin Heidelberg.
Speer, R., Chin, J., & Havasi, C. (2017). ConceptNet 5.5: An Open Multilingual Graph of General Knowledge. *AAAI*.