A Comparison of Two Paraphrase Models for Taxonomy Augmentation

Vassilis Plachouras, Fabio Petroni, Timothy Nugent, Jochen L. Leidner

the answer company™
THOMSON REUTERS®

Research Questions

Taxonomies are resources for organizing knowledge.

Taxonomies are used in a wide range of tasks such as document classification, search and natural language understanding.

However, developing taxonomies is time consuming.

We investigate the automatic augmentation of an existing taxonomy using generative paraphrasing.

RQ1 Can the models generate high quality paraphrases for automatically augmenting a taxonomy?

RQ2 How much does the coverage of the taxonomy increase?

RQ3 Which model is best for generating paraphrases?

Paraphrase Generation Models

We approach the task of generating phrasal paraphrases as monolingual translation and train two state-of-the-art models:

1. **Moses** [P. Koehn et al.(2007)];
2. **seq2seq** [Bahdanau et al.(2015)].

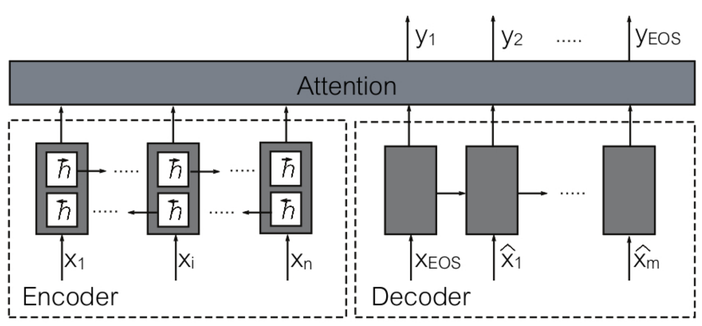
Moses

Moses is an open-source statistical machine translation model. We focus on phrase-based translation and a tri-gram language model learned from the set of target paraphrases.

`http://www.statmt.org/moses`

Attention-based seq2seq

The attention-based seq2seq model consists of a bi-directional LSTM encoder and an LSTM decoder which uses an attention mechanism to learn which input words are the most important for each output word.



Training Corpus And BLEU Score

For training the generation models, we used the Paraphrase Database (PPDB 2.0) corpus.
`http://paraphrase.org`

Paraphrase.org

Model	BLEU
Moses	0.4098
seq2seq	0.3156

BLEU is calculated on tokenized data using the implementation provided in the nltk framework (`http://www.nltk.org`).
Moses is better than the seq2seq model on PPDB 2.0.

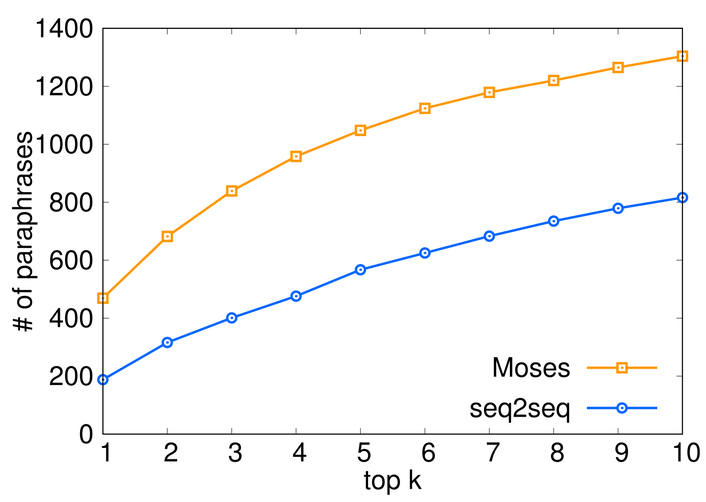
Augmenting The Taxonomy Of Risks

The risk taxonomy consists of 2,824 high quality risk terms. For each term, we apply the two paraphrase generation models to obtain the top 10 paraphrased risk terms.
Examples of paraphrases generated by Moses and seq2seq:

Risk Term	Moses	seq2seq
wind-blown debris	wind-blown rubble	buildings
unexpected entry of competitors	unpredicted entrance of competitors	accident
trafficked people	trafficking in persons	victims of human trafficking
demolished	razed	demolition
committed fraud	fraud committed	the fight against fraud
genetically modified food	gm food	genetically engineered

Stats

Number of top-k generated paraphrases already in the list of risk phrases.



Top-1 Generated Paraphreases

We have manually annotated the top-1 generated paraphrases that were not already in the original risk taxonomy.

Model	Valid	Noisy	Invalid
Moses	1,337	337	327
seq2seq	419	175	2,042

- *valid* when it can directly replace the original risk term;
- *noisy* when the meaning of the paraphrase is close to the meaning of the original term or the paraphrase has additional terms;
- *invalid* when the paraphrase is not suitable for substituting the original risk term.

Coverage Experiment And Research Answers

Number of sentences from the Reuters News Archive (14 million news articles) matching at least one of the generated valid or noisy risk paraphrases, which were not already matched by a risk phrase in the original taxonomy.

Model	Valid	Noisy
Moses	5,197,781	1,868,734
seq2seq	1,751,745	749,886

Paraphrase generation models can expand an existing taxonomy of risk terms with high quality phrases (67% valid) (**RQ1**)

This has led to an increase of the coverage of the taxonomy by 22% (original risk phrases match 23,110,506 sentences) (**RQ2**)

The experimental results also demonstrate that Moses outperforms the neural network-based model in this setting (**RQ3**)

Lexical Diversity

Lexical diversity of generated valid and noisy paraphrases in terms of fraction of tokens that are not in the original risk phrase.

Model	diversity (valid)	diversity (noisy)
Moses	0.5455 (1,337)	0.3952 (337)
seq2seq	0.6991 (419)	0.6969 (175)

The seq2seq model results in higher lexical diversity than Moses for both the valid and noisy paraphrases.

References

[Bahdanau et al.(2015)] Bahdanau et al. 2015. Neural machine translation by jointly learning to align and translate. In *Proceedings of the International Conference on Learning Representations*.

[P. Koehn et al.(2007)] P. Koehn et al. 2007. Moses: Open source toolkit for statistical machine translation. In *Proceedings of the Demo Sessions of the 45th Annual Meeting of the Association for Computational Linguistics*.