Natural Language Processing 365

Day 186: NLP Papers Summary – Contextualizing Citations For Scientific Summarization Using Word Embeddings And Domain Knowledge

Objective and Contribution

Proposed an unsupervised method that uses word embeddings and domain knowledge to extract context from reference paper and achieved SOTA results. Citation contextualisation can improve citation text and downstream applications such as scientific summarisation.

WHAT IS CITATION TEXT?

Citation text are texts that highlight contributions of the referenced paper. Existing work as shown that citation text are useful for many downstream tasks such as search and summarisation. The problem with citation text is that it often lacks context such as methodology, assumptions or conditions for the obtained results.

Contextualising Citations

Our approach extends language models by incorporating word embeddings and domain ontology. The purpose of the language model is to rank a document (reference spans) according to the query (the citation text). We used embeddings to score documents that are semantically similar to the chosen query. We also used domain ontologies to capture information that might not be included in the embeddings. The domain ontologies can extend the embedding based language models as follows:

1. *Retrofitting*. We used MESH and Protein Ontologies (PO) to modify word embeddings and pull synonymous words closer together
2. *Interpolating*. We directly modified the language model and incorporate the domain knowledge

Experiments

We used TAC 2014 Biomedical Summarisation dataset which consists of 220 scientific biomedical journal articles and 313 citation texts (which include relevant contexts). Our baseline models are:

1. BM25
2. Vector Space Model
3. Dual Embedding Space Model
4. Language Modelling with LDA smoothing

We compare the performance of training using the Wikipedia embeddings and biomedical embeddings. In addition, we also compare the performance of biomedical embeddings with retrofitting and with interpolating domain knowledge. In terms of evaluation metrics on the quality of extracted citation contexts (intrinsic evaluation), we have evaluation s for assessing the quality of retrieved contexts for each citation from multiple aspects:

1. Character offset overlaps between retrieved contexts and human annotations in ter. ^ of precision, recall, and F-score

2. nDCG scores where we treat any partial overlaps with the goal standard as correct context

3. ROUGE-N scores to measure content similarity between retrieved contexts and gold standard

4. Character precision at K

Results

INTRINSIC EVALUATION

The results are displayed in the table 1 below. As shown, our model of biomedical embeddings with domain knowledge interpolation achieved the best performance in most of the evaluation metrics. This indicates that our models are effective in many different aspects. We found that general embedding does not provide much advantage when compared to the best performing baseline model. However, when we used the domain-specific embeddings, we observed a 10% c-F improvement over the best performing baseline model. This is as expected as the biomedical embeddings are better at capturing the word relations in biomedical context. This is illustrated by table 2 where we showcase top relevant words for the word "expression" provided by the two types of embeddings. The last two rows of table 1 showcase the benefit of including domain ontologies and we found that interpolation yielded a stronger improvement than retrofitting.

Table 1: Results on TAC 2014 dataset. c-P, c-R, c-F: character offset Precision, Recall and F-1 scores; RG: ROUGE; c-P@K: character offset precision at K. †shows statistical significant improvement over the best baseline performance (two-tailed t-test, $p<0.05$). Values are percentages.

Method	c-P	c-R	c-F	nDCG	RG-1	RG-2	RG-3	c-P@1	c-P@5
VSM [4]	20.5	24.7	21.2	48.1	49.5	26.4	20.0	31.9	26.1
BM25	19.5	18.6	17.8	38.1	43.6	23.2	16.3	25.5	24.2
DESM [12]	20.3	23.8	22.3	45.6	50.3	26.2	20.6	32.5	26.5
LMD-LDA [10]	22.6	24.8	22.3	46.0	48.3	26.4	20.1	31.4	27.7
QR [4]	22.2	29.4	23.8	49.8	50.6	27.2	21.8	37.7	28.1
WE_Wiki	21.8	28.5	23.2	†52.8	50.0	26.9	20.9	36.5	29.9
WE_Bio	23.9	†31.2	†25.5	†57.1	51.9	†29.2	†23.1	†46.2	†34.1
WE_Bio+rtrft	†24.8	†33.6	†26.4	†58.3	52.4	†30.7	†24.0	†55.5	†34.9

Table 2: Top relevant words to the word "expression" according to embeddings trained on Wikipedia vs. Genomics.

General (Wikipedia)	Biomedical (Genomics)
interpretation	upregulation
sense	mrna
emotion	induction
function	protein
intension	abundance
manifestation	gene
expressive	downregulation

Our models also correlate well with human annotations. As shown in table 3 below, when the human precision is high (upper quartiles), our system performs better in c-F metric with higher confidence (lower standard deviation).

EXTERNAL EVALUATION

Citation-based summarisation can effectively capture different contributions and aspects of a research paper by using the citation texts. Here, we compare the summarisation quality between no contextualisation and our proposed contextualisation approaches. We have multiple baseline models and our evaluation metric is ROUGE scores:

1. LexRank
2. LSA-based
3. SumBasic
4. KL-Divergence

The results are displayed below. The "No context" showcase the performance of all summarisation approach on solely the citations and without any context. All the others are different summarisation algorithms on citation text with our contextualisation mets. Our

best performing model is biomedical embeddings with domain knowledge interpolation, ⌃ ch significantly improves the quality of summaries in terms of ROUGE scores. This showcase that the intrinsic quality of citation contextualisation has a direct impact on the end-quality of generated summaries.

Source: https://arxiv.org/pdf/1705.08063.pdf

Ryan

Data Scientist

Previous Post

Next Post

Day 185: NLP Papers Summary -

Day 187: Learn NLP With Me –

A Discourse-Aware
Attention Model for
Abstractive
Summarization of
Long Documents

Embeddings of
Language,
Knowledge
Representation,
and Reasoning

Data Science   Natural Language Processing   NLP Papers Summary

Day 365: NLP Papers Summary – A Survey on Knowledge Graph Embedding

Ryan
30th December 2020

Data Science   Implementation   Natural Language Processing

Day 364: Ryan's PhD Journey – OpenKE-PyTorch Library Analysis + code snippets for 11 KE models

Ryan
29th December 2020

Data Science   Ryan's PhD Journey

Day 363: Ryan's PhD Journey – Literature Review – Knowledge Acquisition – 1st Passes XIV

Ryan
28th December 2020