Article Re-ranking by Memory-Enhanced Key Sentence Matching for Detecting Previously Fact-Checked Claims

Qiang Sheng1,2, Juan Cao1,2, Xueyao Zhang1,2, Xirong Li3, Lei Zhong1,2
1Key Lab of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences
2University of Chinese Academy of Sciences 3Key Lab of Data Engineering and Knowledge Engineering, Renmin University of China

Introduction

- Background: False claims that have been previously fact-checked can still spread on social media.
- Task: Given a claim, retrieval against the fact-checking article (FC-article) collection to find the corresponding fact-checking articles, if any.
- Formulation: Typically formulated as a two-stage retrieval-based workflow.
- We focus on matching the claim with key sentences in FC-articles for better reranking.

Motivation

Existing works ignore the following characteristics of FC-articles:
- Claims are often quoted in key sentences.
  - (Lexical Information, e.g., S2)
- Templates to introduce or debunk claims are common across articles.
  - (Pattern Information, e.g., S2 and S3)

Our Method

MTM (Memory-Enhanced Transformers for Matching)

Input: The given claim q paired with each sentence s in the candidate FC-article d.

1. ROUGE-guided Transformer (ROT): The claim-sentence (q-s) pairs are fed into ROT to obtain claim-sentence vectors and scores. The ROT is a part of pretrained BERT that was finetuned with ROUGE scores as supervision, which considers both semantic and lexical information in q-s pairs.

2. Pattern Memory Bank (PMB): The PMB stores the pattern vectors, which are initialized with the clustering results of the embedding differences in candidate q-s pairs. The distance between the residual embedding (subtracting q from s) and the nearest pattern vector indicates how possible s contains pattern information.

3. Key Sentence Selection: The claim-sentence scores (from ROT) and the pattern-sentence scores (from PMB) are aggregated to rank each sentence in d. Top k sentences are regarded as d’s proxy.

Step 1: Key Sentence Identification

Step 2: Article Relevance Prediction

Memorized Pattern Vectors Update during Training

- An epoch-wise update.
- For each pattern vector m, we collect all residual embeddings whose nearest pattern vector is m.
- We update m to make it closer to the rightly-predicted samples and far away from those wrongly-predicted ones.

Evaluation

Table 2: Performance on baseline and MTM. Best results are in bold.

Method | Sentence Score | Match | BERT | MBT | MAP | MAPt | MTM
--- | --- | --- | --- | --- | --- | --- | ---
ROUGE | 0.83 | 0.64 | 0.94 | 0.63 | 0.86 | 0.89 | 0.89
BERT | 0.83 | 0.64 | 0.94 | 0.63 | 0.86 | 0.89 | 0.89
Standard BERT (BERT) | 0.71 | 0.56 | 0.63 | 0.55 | 0.74 | 0.69 | 0.62 | 0.7 | 0.7 | 0.72 | 0.66 | 0.48 | 0.49 | 0.7 | 0.7 | 0.71
Sentence MBT | 0.75 | 0.55 | 0.62 | 0.56 | 0.74 | 0.72 | 0.64 | 0.72 | 0.74 | 0.73 | 0.63 | 0.46 | 0.55 | 0.48 | 0.56 | 0.57 | 0.71
PMB | 0.84 | 0.68 | 0.96 | 0.68 | 0.91 | 0.91 | 0.87 | 0.91 | 0.92 | 0.91 | 0.87 | 0.54 | 0.63 | 0.56 | 0.63 | 0.65 | 0.71

MTM outperforms all compared methods on the two datasets (the exception is only the MAP@1 on Twitter).

Visualization of residual embeddings around pattern vectors shows the ability of MTM to mine the fact-checking patterns.

Human evaluation on 370 samples show that MTM can find at least one key sentence in 83% of FC-articles (a) and 73% are at Rank 1 (b), even though the positional distribution of key sentences are scattered throughout the articles (c).

Conclusion

- Method: We propose MTM to select from FC-articles key sentences that introduce or debunk claims, and exploit the selected sentences for estimating the relevance of the FC-articles w.r.t. a given claim.
- Evaluation: Experiments show that MTM outperforms existing methods. Further human evaluation and case studies prove that our model can find key sentences, which can be regarded as explanations.
- Data: We built the first Chinese dataset for fact-checking claim detection with fact-checking articles from diverse sources.