Introduction

- Fact-checking of textual sources needs to effectively extract relevant information from large knowledge bases.
- A large-scale fact-checking task, in which verification of claim and extraction of related evidence are required [Thorne et al, 2018]
- Verification labels: Support, Refute and Not enough information (NEI)

Document Retrieval ($DR_{rernk}$)

- A document retriever that searches the whole Wikipedia to find the relevant documents
- Use TF-IDF to reduce the search space from 5.4M to 100 documents
- Apply re-ranking using a scoring function $f_{\text{rank}}$ that utilizes POS tags (NN, NNS, NN, NNPS, JJ, CD), then select the top 5 documents.

$$r_{\text{claim}} = \frac{\text{POS}_{\text{match}}}{\text{POS}_{\text{claim}}} \\ r_{\text{title}} = \frac{\text{POS}_{\text{match}}}{\text{POS}_{\text{title}}} \\ f_{\text{rank}} = r_{\text{claim}} \times r_{\text{title}} \times \text{tf-idf}$$

Evidence Selection ($DA_{\text{rank}}$)

- A neural ranker that extracts $l$ sentences as evidence candidates for given claim using decomposable attention (DA) model
- Trained using a fake task, which is to classify whether a given sentence is an evidence of a given claim or not.
- $l$ value is selected dynamically based on the output evidence score of $DA_{\text{rank}}$; which is considered as a confidence measure of a given sentence being an evidence. Evidence with the score below fixed threshold value $th$ is eliminated.

Lexical Tagging

- Part-of-speech (POS) and named entity recognition (NER) are used to enhance the performance.
- Helps in keyword extraction for each claim.
- Reduces the out-of-vocabulary (OOV) problems related to name or organization entities, for better generalization.

Task results

<table>
<thead>
<tr>
<th>TF-IDF</th>
<th>$DA_{\text{rank}}$</th>
<th>$DA_{\text{rank}} + \text{NER}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy (%)</td>
<td>MLP $DA_{\text{rte}}$</td>
<td>$DA_{\text{rte}} + \text{NER}$</td>
</tr>
<tr>
<td>63.2</td>
<td>78.4</td>
<td>79.9</td>
</tr>
</tbody>
</table>

De-noising

- Prior modules that can effectively leverage the trade-off between recall and precision (high F1) perform the best
- Since the most important factor is to correctly provide succinct set of evidence for the final RTE module.

We propose a framework that verifies a given claim by extracting a set of evidence from Wikipedia.
We extend an existing pipeline [Thorne et al, 2018] by incorporating lexical tagging and de-noising approaches, and proposing neural ranker.

The Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong