Robust Document Retrieval and Individual Evidence Modeling for Fact Extraction and Verification.

Problem Statement
Given a claim involving one or more entities (mapping to Wikipedia pages), the system must extract textual evidence (sets of sentences from Wikipedia pages) that is relevant to the claim and then using this evidence, it must label the claim as Supported, Refuted or Not Enough Info.

Contributions
Robust document retrieval approach leading to coverage of 94.4% of the claims requiring evidence, compared to 55.30% of the baseline method on shared task development set.

Individual evidence modeling and entailment label prediction using majority voting algorithm.

FEVER score of 49.06 on the blind test set with a leaderboard position of 6 out of 24 teams.

Document Retrieval
Google Custom Search API: retrieve documents relevant to the claim.

Wikipedia Python API: collect top documents for each named entity in the claim. These named entities were extracted using pre-trained bidirectional Language Model (Peters et al., 2017).

Query the Wikipedia API: with all the tokens before the first lower case VP in the dependency tree of the claim.

Example: Finding Dory was directed by X.

Sentence Selection
Bigram TF-IDF binning was used to select the top 5 sentences from the k relevant documents.

ELMo embeddings were used to convert the claim and evidence to vectors keeping only the top 3 (out of 5) in cosine similarity.

All five sentences were returned as predicted evidence but only the top three sentences were used for entailment.

Textual Entailment
Individual Evidence Modeling: train model on each claim-evidence pair rather than evidence concatenation.

For recognizing textual entailment we introduced the model introduced by Conneau et al. (2017). BILSTMs with max-pooling to encode the claim and the evidence.

Algorithm for final predictions: SUPPORTS = S
REFUTES = R
NOT ENOUGH INFO = N
C is a count function

\[
\text{Algorithm 1: Prediction Rule}
\]

\[
\begin{align*}
\text{if } C(S) + 1 & \land C(N) + 2 \text{ then } \text{label} = S \\
\text{if } C(R) = 1 & \land C(N) + 2 \text{ then } \text{label} = R \\
\text{else } & \text{label} = \arg \max (C(S), C(R), C(N))
\end{align*}
\]

Sentence encoder with CLAM
Sentence encoder with EVIDENCE

Conneau et al. (2017)

Conclusion and Future Work
- Challenging task primarily because the annotation requires substantial manual effort.
- Presented an end-to-end pipeline to automate human effort and showed empirically a model that outperforms the baseline by a large margin.
- Provided a thorough error analysis which highlights some of the shortcomings of the models and potentially of the gold annotations.
- Future work:
  - Different approaches for entity linking and disambiguation.
  - Joint models for evidence extraction and textual entailment to reduce error propagation.


