

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

OVERVIEW

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.

Team: COLUMBIANLP.

<https://github.com/tuhinjbcse/FEVER-EMNLP>

METHOD

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 used 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

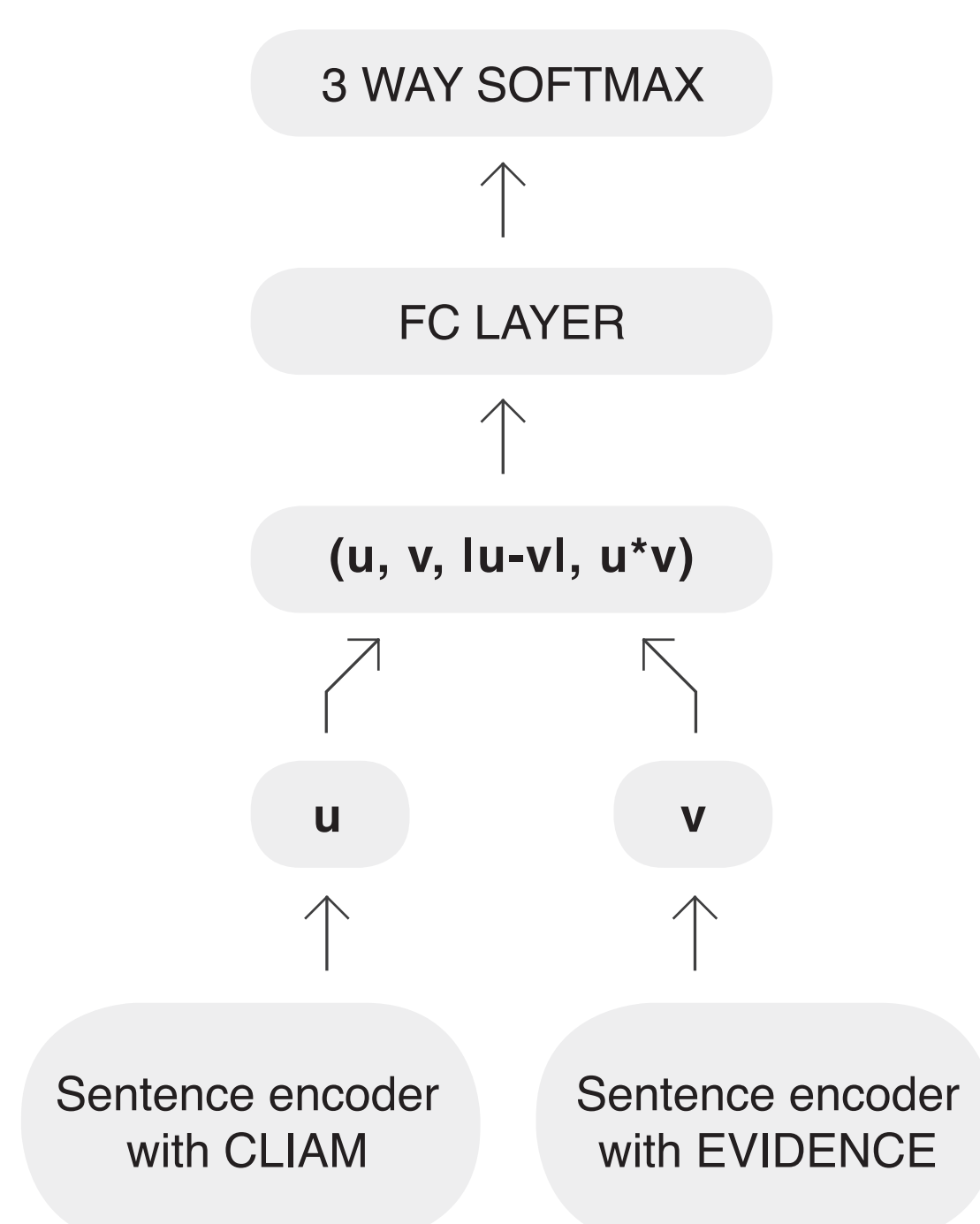
NOT ENOUGH INFO = N

C is a count function

Algorithm 1: Prediction Rule

```

if C(S) = 1 & C(N) = 2 then
    label = S
else if C(R) = 1 & C(N) = 2 then
    label = R
else
    label = arg max (C(S), C(R), C(N))
    
```



Conneau et al. (2017)

Results of Individual Stages

Coverage of claims Supported or Refuted by retrieved documents

METHOD	AVG K	COVERAGE
GOOGLE API	2	79.5%
NER	2	77.1%
DEPENDENCY PARSE	1	80.0%
COMBINED	3	94.4%
THORNE ET AL (2017)	5	55.3%

Evidence recall on both Dev and Test

DATASET	RECALL
SHARED TASK DEV	78.40
BLIND TEST SET	75.89

Entailment accuracy on both Dev and Test

DATASET	ACCURACY
SHARED TASK DEV	58.77
BLIND TEST SET	57.45

End to End Results

Confusion Matrix for Entailment Prediction

	SUPPORTS	NOT ENOUGH INFO	REFUTES
SUPPORTS	4635	1345	686
NOT ENOUGH INFO	2211	3269	1186
REFUTES	1348	3269	3848

Overall FEVER score

DATA	PIPELINE	FEVER
DEV	THORNE ET AL (2018)	31.27
	OURS	50.83
TEST	THORNE ET AL (2018)	27.45
	OURS	49.06

Error Analysis

Claim "Aristotle spent time in Athens", Predicted evidences 'Supports' claim but penalized for not matching the gold evidence.

Better understanding of semantic similarity between SUPPORT and NOT ENOUGH INFO is needed.

Predicted Evidence 1: At seventeen or eighteen years of age, he joined Plato's Academy in Athens and remained there until the age of thirty-seven (c. 347 BC)

Claim: Happiness in Slavery is a gospel song by Nine Inch Nails

Predicted Evidence 2: Shortly after Plato died, Aristotle left Athens and at the request of Philip II of Macedon, tutored Alexander the Great beginning in 343 BC

Evidence: Happiness in Slavery is a song by American industrial rock band Nine Inch Nails from their debut extended play (EP), Broken(1992)

Annotation Errors

Gold Label: Not Enough Info
Correct Label: Supports

Gold Label: Not Enough Info
Correct Label: Refutes

Claim: Natural Born Killers was directed by Oliver Stone.
Evidence: Natural Born Killers is a 1994 American satirical crime film directed by Oliver Stone and starring Woody Harrelson, Juliette Lewis, Robert Downey Jr., Tom Sizemore, and Tommy Lee Jones.

Claim: Anne Rice was born in New Jersey
Evidence: Born in New Orleans, Rice spent much of her early life there before moving to Texas, and later to San Francisco.

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.