OpenTFV: An Open Domain Table-based Fact Verification System

Zihui Gu*, Ruixue Fan*, Xiaoman Zhao, Meihui Zhang, Ju Fan, Xiaoyong Du
[Motivation]

• The prevalence of **misinformation**, both online and offline, has a bad impact on society.
  
  • In view of the spread of COVID-19, the World Health Organization (WHO) claimed that “We’re not just fighting an epidemic, we’re fighting an **infodemic**”.

• An example about **infodemic**:

  "**FAKE NEWS**"  
  "Heavy smokers have a lower COVID-19 hospitalization rate than non-smokers."

  **MORE evidence emerges that smokers are protected from coronavirus**: Italian study finds them FIVE TIMES less likely to end up in hospital (but almost twice as likely to die if they do)

  - Of 441 hospitalised COVID-19 patients in the study, 8% were active smokers
  - But rates are 24% in the Italian population, suggesting some sort of protection
  - However, more smokers succumbed to the disease - half of smokers compared to 35% of non-smokers
  - It comes after a UCL review of 28 studies, including from Europe, which found smokers were less common than expected among COVID-19 patients
  - Governments suggest quitting the habit for precaution despite lack of evidence
  - Here’s how to help people impacted by Covid-19

  Researchers have uncovered more evidence that smokers could be protected from the deadly coronavirus

We can find evidence from statistical tables in relevant papers for verification!

If you search for keywords “COVID-19” and “smoker” on Google Scholar...

You’ll get about a hundred papers!

But there are problems...

➢ **Time-consuming**: tables consist of massive amounts of data.
➢ **Complex**: the process of fact verification may involve mathematical operations.

You can finally get some relevant statistical tables.
[Related Work]

• Recently, the importance of tabular data has brought a new demand for **automatic table-based fact verification** that verifies if a claim is supported or refuted by tables.

• Most of the existing studies on table-based fact verification are in a **closed-domain** setting. *(Not realistic)*
  - The table relevant to a claim is given during verification. *[Chen et al. ICLR’19]*
  - The schema of tables is known in advance. *[Karagiannis et al. VLDB’20]*

• Some very recent work considers the open domain setting, it utilizes the traditional TF-IDF method and has limitations to retrieve **semantically relevant tables** for claims. *[Schlichtkrull et al. ACL’21]*

---

We introduce OpenTFV, an open domain table-based fact verification system.

We formally define **open domain table-based fact verification** problem as follows:

- Given a claim $s_i$ and a corpus of $M$ tables $D = \{T_i\}_{i=1}^M$.
- Our goal is to retrieve the top-$k$ relevant tables $\tilde{D} = \{T_i\}_{i=1}^K \subset D$ and determine whether the claim $s_i$ is refuted(0) or entailed(1) by each relevant table $T_i \in \tilde{D}$.
• There are three challenges to do table-based fact verification in an open domain setting:

I. Difficulty in **finding** relevant tables from a large table corpus.

II. Difficulty in **verifying** if a relevant table supports or refutes a claim.

III. Difficulty in **interpreting** why a claim is supported or refuted by a table.
Table Data Representation encodes each table in the corpus into a vector-based representation for both table retrieval and fact verification.

- Structure-aware pre-trained LM.
- Table serialization.
- Intermediate pre-training
Table Retrieval aims to retrieve relevant tables efficiently and accurately.

- **Keyword matching** obtains top-k candidate tables with high recall.
- **Semantic-based re-ranking** refines the ranking of the candidate tables.
Fact Verification encodes the serialized table and the claim and produces a verification result (ENTAILED or REFUTED).

- **Program-based analysis** identifies potential operations (e.g. compare) in the claim.
- **Attention-based fine-tuning** encodes a target cell with a specific column and row in a table.
- **Data augmentation** method to generate more claims from the table corpus via a context-free grammar.
Interpretation Generation explains the fact verification process via generating natural language interpretations.

- Template-based method:

  Filter_eq

  Locate column [Parameter] with [value] : row [index]
[Demonstration]

- We demonstrate OpenTFV in two representative scenarios:

  I. **COVID-19 claims fact verification based on academic tables.** We allow the users to input COVID-19 related claims and demonstrate how OpenTFV supports open domain fact verification over a large corpus of academic tables extracted from a COVID-19 Open Research Dataset.

  II. **General fact verification based on Wiki-tables.** Apart from academic tables, we also deploy OpenTFV on top of Wiki-tables extracted from Wikipedia.
OpenTFV is a user-friendly system that supports open domain table-based fact verification.

- Open domain table retrieval
- Interpretation of verification process
- High accuracy verification
- User friendly interface
Thank you!