Generation of Training Examples for Tabular Natural Language Inference

Jean-Flavien Bussotti, Paolo Papotti

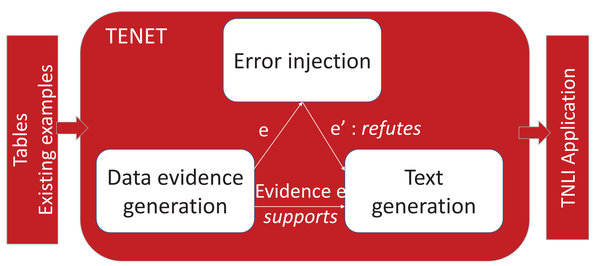
Tabular data is an important source of information in tabular NL inference, (TNLI) such as computational fact-checking tasks. Given tables, existing approaches to generate training data for fact-checking models are either based on expensive human annotations or on methods that produce simple examples. We propose a system, *Tenet*, for the automatic generation of training examples given only the tables as input.

Pipeline

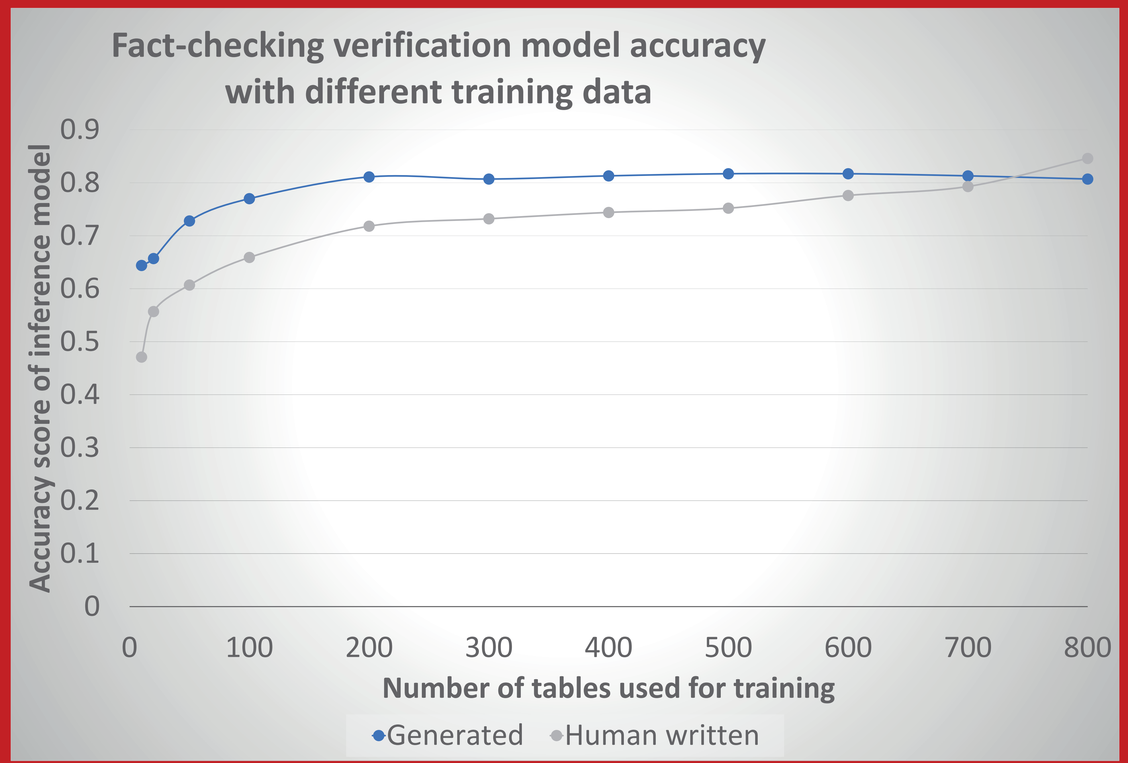
1. Input : Wikipedia table
   Output : generated data for **supports** claim
   - Warm approach: inspired from humans' annotations, using SQL
   - Cold approach : random selection
2. Injecting errors : data for **refutes** claim
   - shuffle evidence columns
   - add (using PLM)/ remove additional fake rows
3. Create textual hypothesis for the tabular table with PLMs
4. Run fact inference model training on the newly created dataset

Results

- Tests on 3 corpora from the TNLI literature + 2 crafted by us on more complex tables
- Warm approach > Cold on aggregate operation in claims + relying on the evidence in tables
- Tenet generates human-like examples,
  - ➢ training of inference algorithms with results comparable to those obtained with manually-written examples



Generating data with appropriate methods can provide **as good quality training data as humans** but costing **less time and money**.



Claim/hypothesis :
"Lindfield railway station has 3 bus routes, in which the first platform services routes to Emu plains and Hornbys and the third platform services routes to Berowra and Gordon"
Label :
Supports
Table & Evidence :

Platform	Line	Stopping pattern	Notes
1	T1	services to Penrith, Emu Plains & Richmond via Central ★	[4]
	T9	services to Hornsby via Strathfield ★	
2	T1	terminating services to/from Penrith & Richmond	
3	T1	services to Gordon, Hornsby & Berowra ★	
	T9	services to Gordon	

Name	Age	City
Mike	47	Nice
John	22	Antibes

Original table

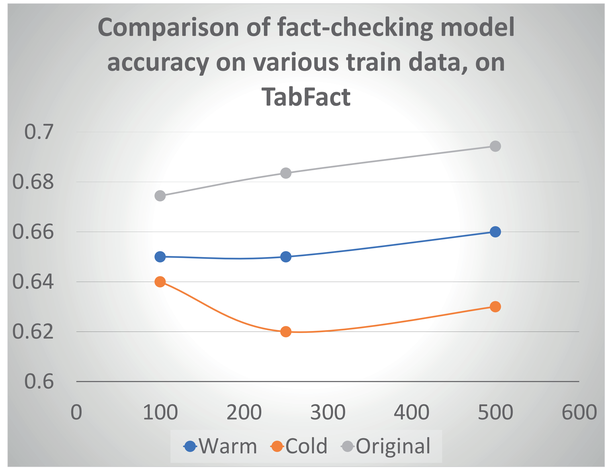
Correct

Name	City
John	Antibes

Incorrect

Name	City
John	Nice
Laure	Cannes

Test set	Warm	Cold	Original
Out of domain	0.84	0.80	0.77
Swapped	0.65	0.65	0.64



EURECOM
Sophia Antipolis