LIPI at the NTCIR-16 FinNum-3 task: Ensembling transformer based models to detect in-claim numerals in financial conversations

•••

Fidelity Investments, Jadavpur University

Sohom Ghosh

Sudip Kumar Naskar

Overview

- Introduction
  - Understanding in-claim versus out-of-claim numerals
- Motivation
  - Help investors to identify out-of-claim numerals so that they do not get allured by in-claim numerals.
- Problem Statement
  - Determine if numerals in financial texts are out-of-claim & in-claim

Dataset

- NTCIR-16 FinNum-3: Investor's and Manager's Fine-grained Claim Detection (English dataset)
- Training: 8,337 records (In-claim: 1,039)
- Validation: 1,191 (In-claim: 114)

Each labelled instance consists of:

- Financial Text
- Numeral (present within the text)
- Starting position of the numeral
- Ending position of the numeral
- Category of the numeral
- Label (In-claim or Out-of-claim)
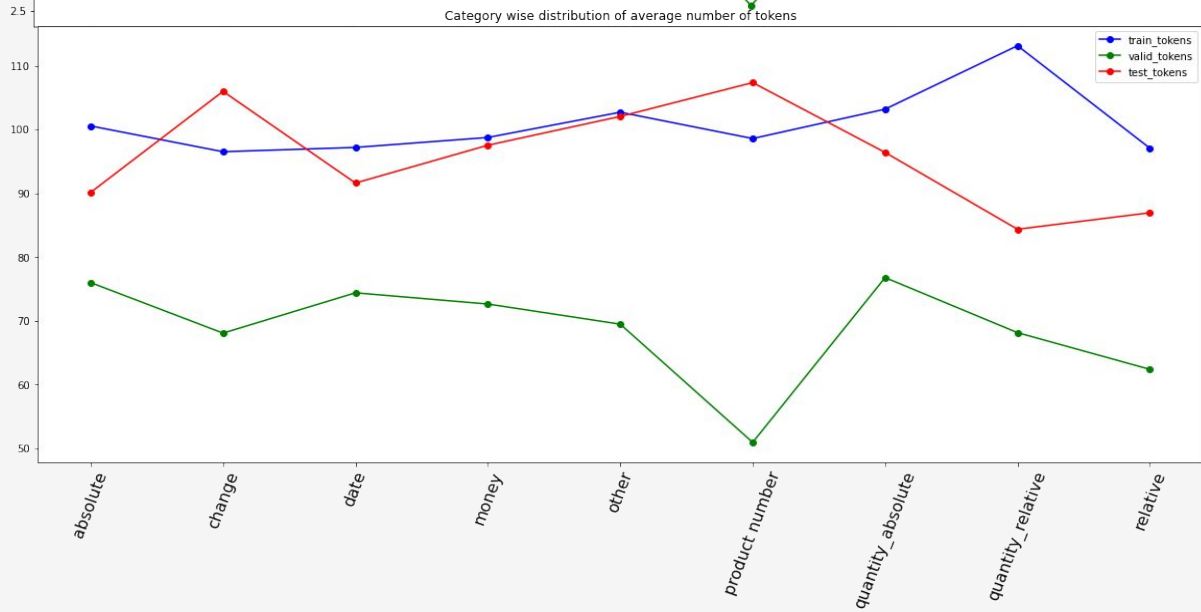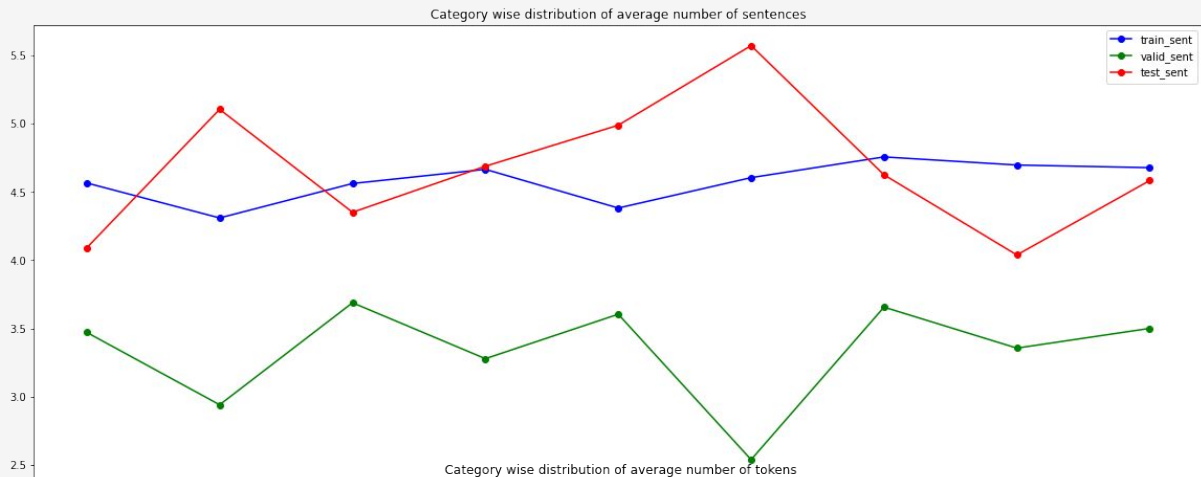
Methodology
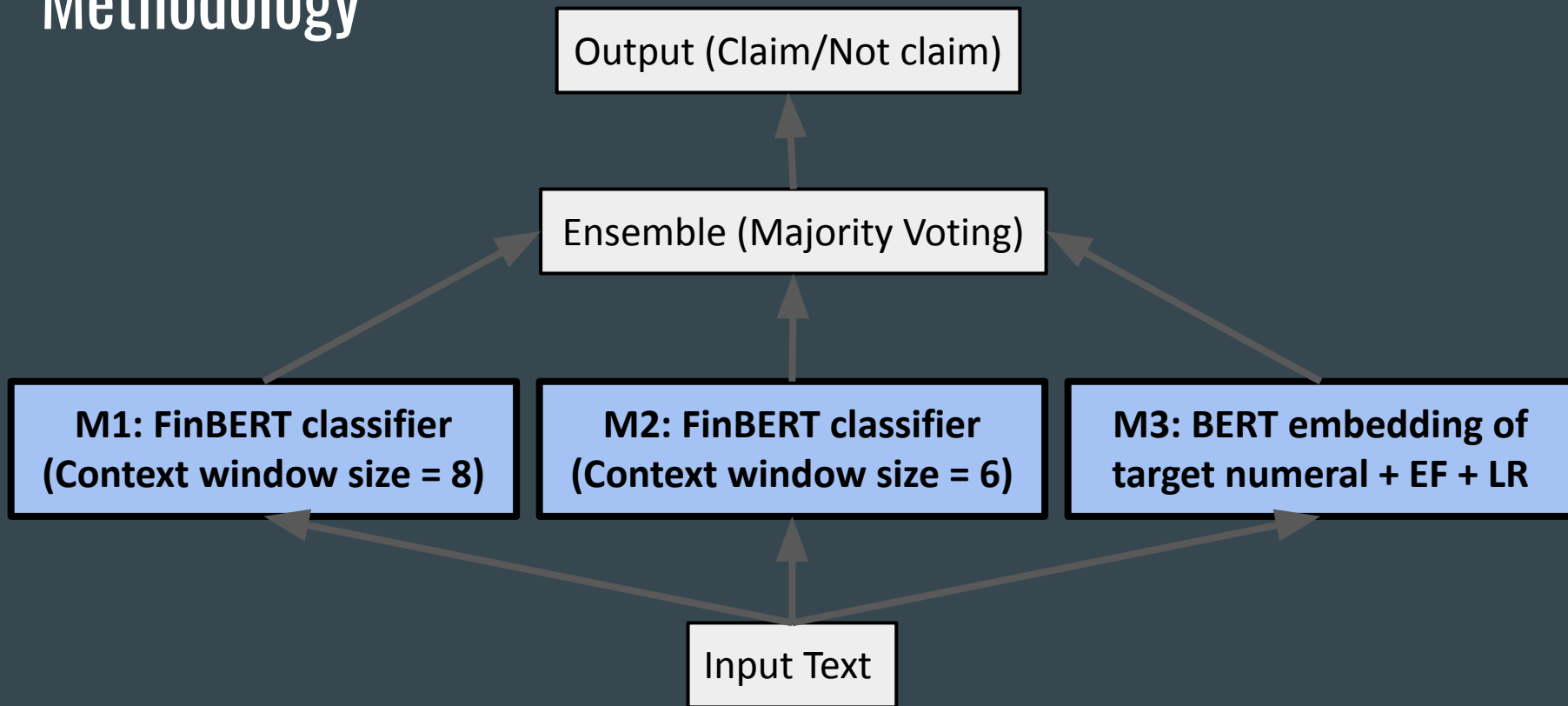
Results & Discussions

submission	Validation		Test	
	Micro-F1	Macro-F1	Micro-F1	Macro-F1
LIPI_1	94.45%	85.84%	95.09%	82.82%
LIPI_2	**95.80%**	**88.15%**	95.17%	81.33%
LIPI_3	94.79%	86.71%	**95.59%**	**84.73%**

Model	Macro-F1	Micro-F1
M3 (only)	0.8318	0.6646
M3 (-EF)	0.8238	0.7990
FinBERT classifier (CW=5)	0.8318	0.7250
FinBERT classifier (CW=6)	0.8439	0.6603
FinBERT classifier (CW=7)	0.8381	0.8244
FinBERT classifier (CW=8) (**LIPI_1**)	0.8585	0.6345
FinBERT classifier (CW=9)	0.8247	0.8262
Ensemble (LIPI_3)	**0.8671**	**0.9479**

Conclusions & Future Works

Conclusions

- We ranked 9th and 10th out of 16 submissions in terms of micro and macro F1 scores respectively
- Developed and open sourced the tools "FiNCAT"- Financial Numeral Claim Analysis Tool and "FiNCAT-2" (https://github.com/sohomghosh/FiNCAT_Financial_Numeral_Claim_Analysis_Tool/)

Future Works

- Take the numeral to be evaluated as an input from the user
- Improving model performance
- Web browser extension development

Thank You

Any Questions