Multi-Entity Aware Sentiment Analysis for Financial Headlines

W266 Final Project

Noah Prozan & Yifan Wang

Background

- Motivation
  - Efficient Market Hypothesis
  - Trading Efficiency
- Related topics
  - Named-Entity Recognition
  - Sentiment Classification
  - Aspect Based Sentiment Analysis (ASBA)
- Challenge:
  - Multiple entities in one headline
  - Conflicting sentiments

Data

SEntFin 1.0 (Main Data)

- 10753 news headlines of Indian stock market
- Annotated: entities + sentiments
- 2847 headlines w/ more than 2 entities
- 1233 headlines w/ conflicting sentiments
- Columns:
  - Title [String]: News headlines
  - Decisions [Dictionary]:
    - {entity1: sent1, entity2: sent2, ...}
    - Sentiment categories: negative, neutral, positive
  - Words

SemEval, FiQa (Secondary Data)

- SemEval: 1647 news headlines, FIQA: news 438 headlines
- Other purposes: identifying industries, microblogs, etc.
  - 1296 usable sentences
- Annotated: entities & scoring sentiments
- JSON format
- Scoring in -1 (very negative) to 1 (very positive)

Limitations of Existing Study

SEntFin Paper:

- **Unscalable process:**
  - Manually labeled entities
- **Inappropriate scoring:**
  - Based on **isolated** entity+sentiment labeling
  - No consideration for a headline as a whole
  - **Limited business value**
  - **Inflated** accuracy
- **Indian stocks**
  - Emerging market: **market inefficiencies**
  - Real-world monetization

Model:
Baseline

A Unified T5 Model

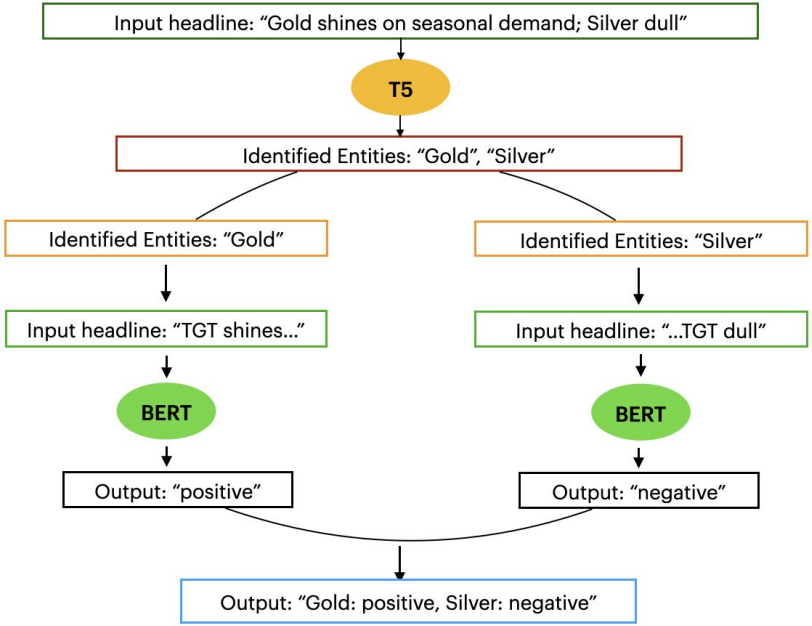
Literature: Towards Generative Aspect-Based Sentiment Analysis

Justification: Combines tasks in one step, conveniently adapted

Input: FY16 not a good year for large drug makers; prefer midcap pharma stocks

Target output: midcap pharma stocks: positive, large drug makers: negative

Model
A Two-Step Approach



Step-1: T5 for NER

- **Example Input:** Gold shines on seasonal demand; Silver dull
- **Example Target Output:** Gold, Silver
- **Best single-step result:** 80.8%

Then: Break down tasks

Step-2: RoBERTa for Sequence Classification

- **Example Input:** TGT shines on seasonal demand; Silver dull
- **Example Target Output:** positive
- **Best-performing model:** RoBERTa
- **Best single-step result:** 87.5%

Finally: Collate results to match target format

Model:
BIO-Token-Classification

BIO Tagging System

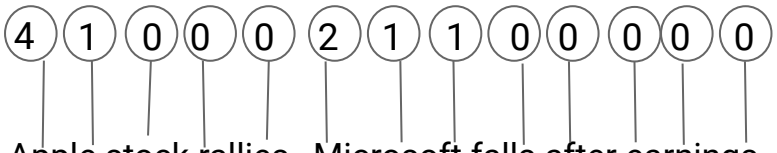
4, if token is “B” beginning of entity and sentiment positive

3, if token is “B” beginning of target and neutral

2, if token is “B” beginning of target and negative

1, if token is “I” , in the target, but not beginning

0, if token is “O”



EX: Apple stock rallies , Microsoft falls after earnings

General Methodology

Step-1: BIO NER Tagging System

Step-2: Conventional BERT scheme using finBERT & RoBERTa implemented Via Hugging Face

Step-3: Dense Layer/ Dropout

Step-4: Softmax Token Classification output of most likely label

Variations

FinBERT, RoBERTa, RoBERTa with pre-train, RoBERTa-CRF

Results

Model Type	Model	Full-Sentence Accuracy	Adj. Full-Sentence Accuracy
Baseline	*T5 Unified*	*61.01%*	*64.81%*
BIO-Token-Based Classification	FinBERT	65.18%	69.17%
	RoBERTa	67.60%	72.42%
	Pretrain+RoBERTa	68.62%	72.61%
	RoBERTa+CRF	70.84%	74.65%
Two-Step Classification	**T5+RoBERTa**	**72.79%**	**76.23%**
	T5+Pretrain+RoBERTa	71.96%	75.21%

Evaluation Metrics:

- Full-sentence accuracy
  - All tokens correctly labeled
  - Output matches in entirety
- Adjusted full-sentence accuracy
  - First token/word of all entities correctly identified
  - Sentiment for all entities correctly classified

Model Analysis

Example Headline	Model Outputs	Target Outputs	Description
Sterling hits 2-week low vs dollar, as housing fervour cools	{'Sterling': 'negative', 'dollar': 'neutral'}	{'Sterling': 'negative'}	*Wrong NER Tagging*
Gold's longest run in a year ends as US inflation picks up	{'Gold': 'neutral'}	{'Gold': 'negative'}	*Ambiguous sentiment*
Standard Chartered axes 15,000 jobs, raises $5.1 billion in capital	{'Standard Chartered': 'negative'}	{'Standard Chartered': 'positive'}	*Failure to digest finance specific information*
Kwality among cheapest plays in dairy sector: Ashish Maheshwari	{'Kwality': 'negative'}	{'Kwality': 'positive'}	*Failure to understand word in a financial context*
I do not see too much of a downside for Infy: Sandeep Wagle	{'Infy': 'positive'}	{'Infy': 'negative'}	*Target output incorrectly annotated*
A $25 fall in crude is like a $10-billion stimulus for Indian economy, say experts; top stock bets	{'crude': 'positive'}	{'crude': 'neutral'}	*Failure to isolate target-specific sentiment from overall sequence sentiment*
Negative on Tata Motors: Ambareesh Baliga, Way2Wealth Brokers Pvt. Ltd	{'Tata Motors': neutral, 'Way2Wealth: neutral'}	{'Tata Motors': neutral, 'Way2Wealth Brokers' Pvt. Ltd: neutral}	Incorrect capture of entire entity

Key talking points:

- Idiosyncrasies of financial news
- Difficulty identifying full headline
- Polar bias
  - Pre-trained model under performs
- Underperformance of multi-entities

Future Steps/Learnings

- NER needs an entity linking counterpart in this scenario
  - Need for a fully formed entity database to truly map stock prices
- Clean Data is still one of the (if not THE) most important thing in the ML project
- Multiple Entity ASBA is still quite challenging
  - Limitations of CRF
- Larger dataset required
  - Tweets
  - Beyond Indian stock market
  - Microblog threads
  - Commodities
  - Tagging to industries?
- Business relevance
  - How does positive/negative sentiment affect stock prices?
  - Algorithmic trading method

Bibliography

Arslan, Yusuf, et al. “A Comparison of Pre-Trained Language Models for Multi-Class Text Classification in the Financial Domain.” Companion Proceedings of the Web Conference 2021, 2021, https://doi.org/10.1145/3442442.3451375

SEntFiN 1.0: Entity-aware sentiment analysis for financial news. Journal of the Association for Information Science and Technology, 73(9), 1314– 1335. https://doi-org.libproxy.berkeley.edu/10.1002/asi.24634

Cortis, Keith, et al. “Semeval-2017 Task 5: Fine-Grained Sentiment Analysis on Financial Microblogs and News.” Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017), 2017, https://doi.org/10.18653/v1/s17-2089.

“Fiqa - 2018.” FiQA - 2018, https://sites.google.com/view/fiqa/.