language and translation technology team

Fine-grained implicit sentiment processing of polar economic events

Gilles Jacobs and Véronique Hoste

Gold

- Feasibility check: given gold polar facts classify implicit sentiment?
- Model selection: Hyperband hyperparameter search. [4] Model eval.: Train-dev-test + McNemar's significance test.
- Fine-tune several transformers including several in-domain.
- Add lexicon scores from general domain and finance at clf head.

Holdout results. Precision (P), recall (R), F_1-score (F_1) percentages are macro-averaged. Accuracy (A) with the p-value of McNemar's test w.r.t. best on dev RoBERTa$_{Large}$+econ+general.

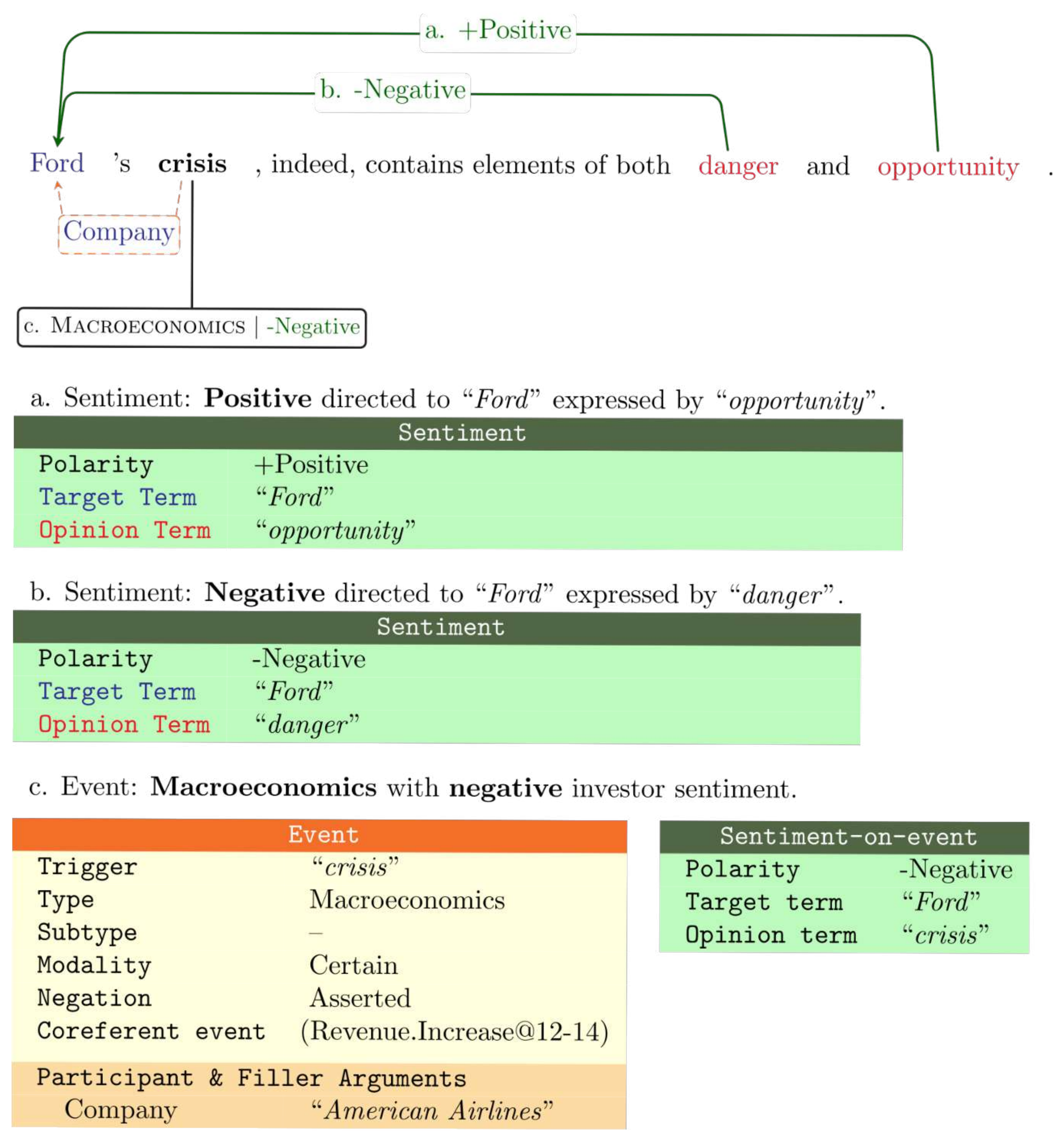
model w lexicons	P	R	F1	A	p
BERT$_{Base}$	57.3	55.4	54.6	68.3	3.6e-10***
+ econ.	62.8	59.0	58.5	71.3	5.6e-05***
+ econ.+general	59.8	57.9	57.3	71.2	3.3e-05***
FinBERT-SST$_{Base}$ [5]	60.8	56.2	56.3	71.2	2.7e-05***
+ econ.	**72.2**	54.0	52.5	73.2	7.1e-03**
+ econ.+general	63.0	57.9	58.4	73.3	2.2e-02*
DeBERTa$_{Base}$ [6]	59.5	58.4	57.9	71.6	7.8e-05***
+ econ.	62.7	57.7	57.7	72.5	1.3e-07***
+ econ.+general	70.4	58.2	58.9	74.9	1.3e-03**
RoBERTa$_{Base}$ [7]	58.4	55.9	54.0	74.5	1.2e-01
+ econ.	63.8	61.5	61.5	74.4	1.4e-01
+ econ.+general	63.5	62.6	62.8	75.0	2.4e-01
BERT$_{Large}$	58.8	56.2	56.0	73.2	9.5e-03**
+ econ.	61.8	61.6	61.5	72.8	4.3e-03**
+ econ.+general	62.0	58.0	58.1	74.0	5.5e-02
RoBERTa$_{Large}$ [7]	65.8	**63.0**	**63.2**	**77.5**	1.1e-01
+ econ.	63.8	62.6	63.0	75.4	5.8e-01
+ econ.+general	61.9	58.8	58.3	75.9	-

- Adding lexicon helps marginally: ↑P but overfit
- RoBERTa is robust, base > many large.

→ 78% Accuracy: polar fact polarity classification is **hard but feasible**.

SENTiVENT Dataset

- **SENTiVENT Dataset**: 286 economic news articles annotated with: ACE/ERE event structures + target-sentiment expressions. → 9500 target + opinions [1, 2]
- **Polar facts**: connotational implicit sentiment of events, facts [3] → challenge in SA
  - Combine events + sentiment for targeted polar fact processing: event participants = targets.
- Applications: market analysis, trading strats, event studies …



- **Investor sentiment**: *"Implicit and explicit text that expresses or affects an investor's attitude towards an economic actor."*

Bullish
Bearish
Neutral =

Clause

- Realistic task: classify **polar clauses:** OpenIE clause extractor + dependency rule fall-back to split original sentences in clauses:

"Passenger count PAX grew 0.8 %, while it declined 0.7 % on a year-to-date basis ."

→ [NEG] *"it declined 0.7 % on a year-to-date basis."*

→ [POS] *"passenger count PAX grew 0.8 %,"*

→ 57% macro-F1, 66% Acc for RoBERTa

- Adding lexicons ↑P, ↓R
- Confusion "none" vs "neutral".

model w lexicons	P	R	F1	A	p
FinBERT-SST$_{Base}$ [5]	53.7	53.9	52.2	60.0	$< .001^{***}$
+ econ.	53.6	54.8	53.4	60.2	$< .001^{***}$
+ econ.+general	52.5	53.1	52.0	60.4	$< .001^{***}$
BERT$_{Large}$	54.8	55.8	54.6	62.0	.018*
+ econ.	56.4	57.1	55.6	62.8	.086
+ econ.+general	55.1	52.3	50.3	61.3	.002**
DeBERTa$_{Base}$ [6]	53.8	53.7	52.5	63.3	.214
+ econ.	56.7	56.0	55.2	63.7	.34
+ econ.+general	52.9	52.4	50.1	64.0	.493
RoBERTa$_{Base}$ [7]	57.8	55.3	55.1	**66.2**	.26
+ econ.	56.2	56.6	55.7	63.7	.373
+ econ.+general	57.6	56.0	54.2	65.5	.541
RoBERTa$_{Large}$ [7]	56.7	**57.8**	56.7	64.8	.944
+ econ.	**59.1**	57.6	**56.8**	64.8	-
+ econ.+general	56.0	55.6	54.6	63.5	.232

Holdout results for clause-based implicit polar fact classification. Precision (P), recall (R), F_1-score (F_1) percentages are macro-averaged. Accuracy (A) with the p-value of McNemar's significance w.r.t. best dev (RoBERTa$_{Large}$+econ.).

Fine

- Fine-grained token-level: Target-opinion-polarity triplet extraction
  (*Ford's share price*, *weakness*, *NEG*)
  (*Ford's share price*, *surpass*, *NEG*)
  (*Tesla*, *surpass*, *POS*)
- Apply end-to-end SotA model "Grid Tagging Scheme for ABSA" (GTS) [9]

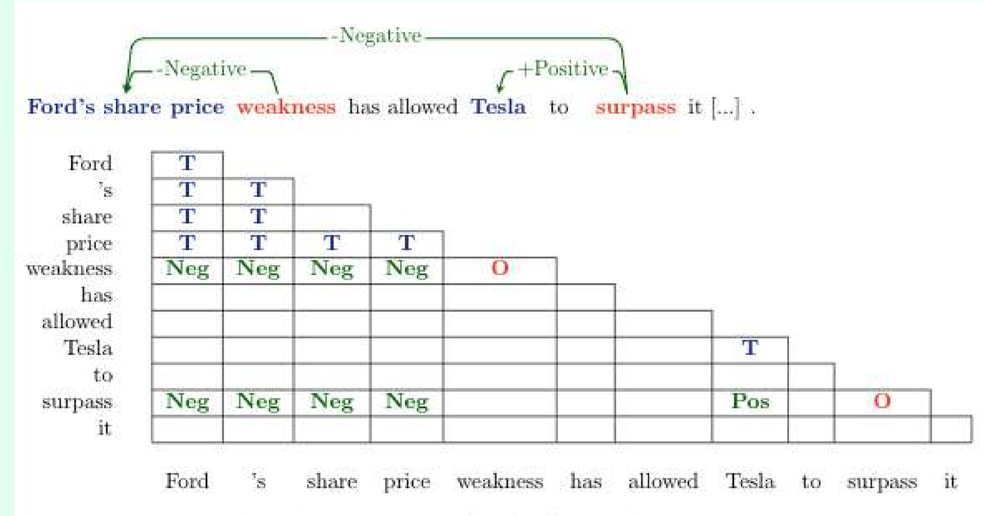


Figure 2: A grid tagging example for the OTE task.

Table 1: Precision (P), Recall (R), and F_1-scores on Opinion term (sentiment token span) extraction, Aspect term (target token span), and combined Triplet extraction (opinion term, aspect term, sentiment polarity).

Dataset + Model	Aspect term			Opinion term			Triplet		
	P	R	F_1	P	R	F_1	P	R	F_1
SENTiVENT Roberta$_{Base}$ [7]	57.9	**60.6**	**59.2**	41.3	**39.12**	40.2	**26.4**	**20.8**	**23.2**
SENTiVENT Roberta$_{Large}$ [7]	**59.0**	57.7	58.4	**43.6**	39.0	**41.2**	24.4	19.2	21.5
SENTiVENT FinRoBERTa$_{Base}$ [8]	31.2	44.1	36.5	20.2	24.5	22.1	8.6	8.1	8.3
Reviews [11] Roberta$_{Base}$ [7]	**85.8**	88.0	86.9	**87.6**	88.0	**87.8**	**75.0**	74.1	74.5
Reviews [11] Roberta$_{Large}$ [7]	84.7	**89.3**	**86.9**	86.6	**88.3**	87.5	74.8	**74.4**	**74.6**

- **Implicit much harder than explicit** :
  - "Closed" economic domain?
  - Beyond lexical methods
  - Polar facts remain a challenge.

→ SENTiVENT ideal resource for pushing polar fact processing

1. Jacobs, G., & Hoste, V. (2020). Extracting Fine-Grained Economic Events from Business News. In Proceedings of the 1st Joint Workshop on Financial Narrative Processing and MultiLing Financial Summarisation (pp. 235-245).
2. Jacobs, G., & Hoste, V. (Forthc). SENTiVENT: enabling supervised information extraction of company-specific events in economic and financial news. Language Resources and Evaluation.
3. Toprak, C., Jakob, N., & Gurevych, I. (2010, July). Sentence and expression level annotation of opinions in user-generated discourse. In Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (pp. 575-584).
4. Li, L., Jamieson, K., DeSalvo, G., Rostamizadeh, A., & Talwalkar, A. (2017). Hyperband: A novel bandit-based approach to hyperparameter optimization. The Journal of Machine Learning Research, 18(1), 6765-6816.
5. Araci, D. (2019). Finbert: Financial sentiment analysis with pre-trained language models. arXiv preprint arXiv:1908.10063.
6. He, P., Liu, X., Gao, J., & Chen, W. (2020). Deberta: Decoding-enhanced bert with disentangled attention. arXiv preprint arXiv:2006.03654.
7. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692.
8. https://huggingface.co/sm6342/FinRoberta
9. Wu, Z., Ying, C., Zhao, F., Fan, Z., Dai, X., & Xia, R. (2020, November). Grid Tagging Scheme for End-to-End Fine-grained Opinion Extraction. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings (pp. 2576-2585).
11. Fan, Z., Wu, Z., Dai, X., Huang, S., & Chen, J. (2019, June). Target-oriented opinion words extraction with target-fused neural sequence labeling. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers) (pp. 2509-2518).

GHENT UNIVERSITY