ECONOMIC EVENT DETECTION FOR COMPANY-SPECIFIC NEWS TEXT

Gilles Jacobs gillesm.jacobs@ugent.be
Els Lefever els.lefever@ugent.be
Véronique Hoste veronique.hoste@ugent.be

Language & Translation Technology Team (LT3), Ghent University
ECONLP @ ACL 2018, Melbourne, Australia

OUTLINE

1. Introduction
2. Dataset description
3. Experiments
4. Error analysis
5. Conclusion
6. Limitations & future work

INTRODUCTION

- Task: Event mention detection and typing for economic events in company-specific news.
- Event conceptualization: "*textually reported real-world occurrences, actions, situations involving companies*"

"However, <u>revenues</u> from voice and text <u>fell</u> in the period."
➜ Turn-over event

"So far, free cash flow has been used to finance <u>share buybacks</u> and dividend increases."
➜ Share buyback event

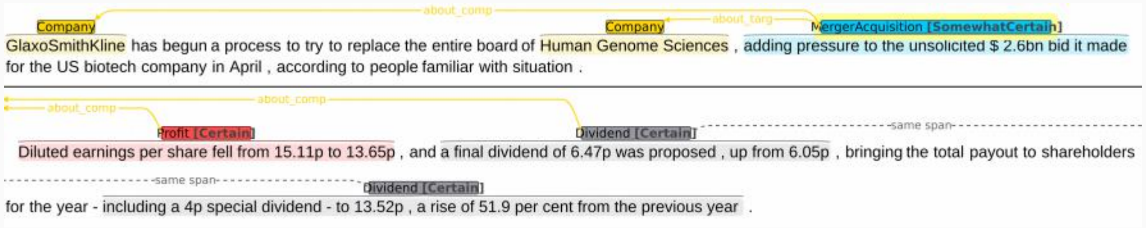
"It will <u>increase the number of Barclays' customers</u> in France by 25 per cent."
➜ Sales Volume event

INTRODUCTION

- Applications:
  - Economics (academic): event studies (MacKinlay, 1997); assessing impact of news events (Boudoukh et al., 2016)
  - Security price prediction, business intelligence, trading strategies, etc.
- Current methods are pattern- or knowledge-based (Feldman et al., 2011; Arendarenko and Kakkonen, 2012; Hogenboom et al., 2013; Du et al., 2016): largely handmade ontologies.
- Supervised, data-driven methods: potential to generalize over lexical variation.
- → requires annotated **gold-standard dataset**.
- Currently no resources exist for supervised event detection in economic domain.
  - In general-domain event detection: large amount of resources (e.g. ACE/ERE (TAC-KBP) (Aguilar et al., 2014)).

DATASET DESCRIPTION

- Our SentiFM dataset enables event mention detection and typing (!= event extraction with argument slots).
- **7 companies** selected for sector diversification.
- **497 news articles** from the Financial Times (2004-2013) in English.
- **2522 event mentions**.
- Annotations:
  - Token-level annotation: Discontinuous and multi-token span.
  - “about_company” relation on each event.



DATASET DESCRIPTION

- Dutch counterpart available (Lefever and Hoste, 2016).
- Validity of the annotation scheme was evaluated on Dutch subset:
  - 78.41% Inter-annotator F1-score.
- Event types:
  - 10 event types, typology constructed iteratively on corpus subsample.
  - Type overlap with StockSonar (Feldman et al., 2011) and SPEED ontology (Hogenboom, 2013).

Event type	**Type ratio**	**# mentions**
BuyRating	9.00%	227
Debt	2.38%	60
Dividend	7.22%	182
MergerAcquisition	10.03%	253
Profit	25.81%	651
QuarterlyResults	10.59%	267
SalesVolume	19.31%	487
ShareRepurchase	2.42%	61
TargetPrice	3.73%	94
Turnover	9.52%	240
total	100%	2522

Event type distribution and mention count.

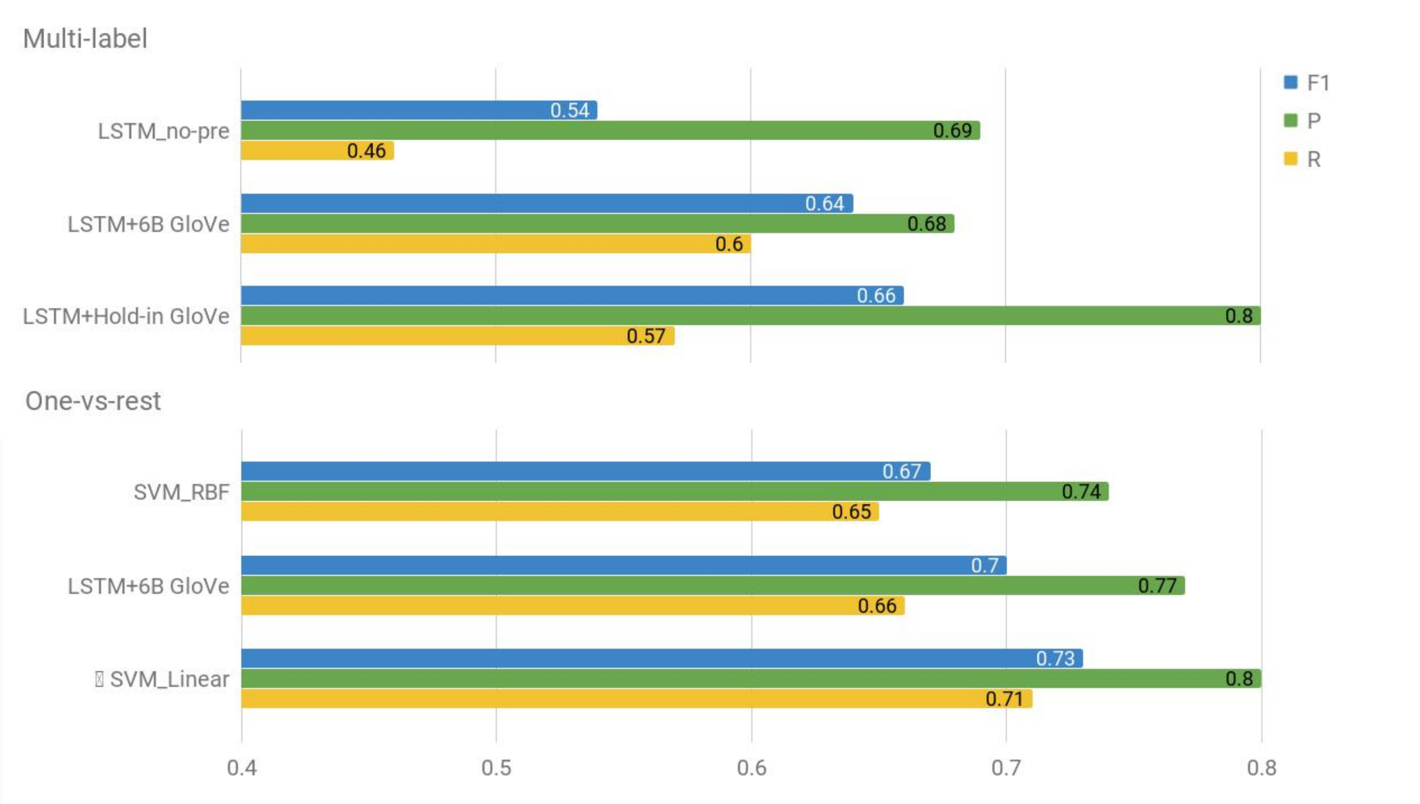
EXPERIMENTS

- Goal: provide a baseline for the dataset.
- Task: Sentence-level event typing, multi-label classification.
- **Feature-engineering + SVM**.
  - Lexical features: bag-of-n-gram (token, char, lemma), special token presence.
  - Syntactic features: PoS and NER-tags.
  - Kernels: linear and RBF.
- **wvec + LSTM**:
  - Pre-trained, hold-in set GloVe.
  - Pre-trained, 6B corpus GloVe.
  - No pre-trained, token-sequence input.
- Pre-trained vectors chosen from multiple candidates by quality evaluation on analogy task.

EXPERIMENTAL SET-UP

- Evaluation: support-weighted macro-F1 on 10% random hold-out test.
- Hyper-parameter optimization:
    - SVM-RBF: 5-fold cross-validation, grid-search.
    - SVM-Linear: No optimization, default LibSVM hyper-parameters.
    - LSTM: 3-fold cross-validation, randomized-search (32 it.).
- SVM: one-vs-rest.
- LSTM: multi-label & one-vs-rest (for best input).

RESULTS: ALL SYSTEMS



RESULTS: BEST SVM & LSTM SYSTEM, SCORES BY TYPE

Event type	Precision	Recall	F_1-score
Linear kernel one-vs-rest			
BuyRating	**0.95**	**0.91**	**0.93**
Debt	**0.50**	**1.00**	**0.67**
Dividend	**0.62**	**0.73**	**0.67**
MergerAcquisition	**0.56**	**0.40**	**0.47**
Profit	0.75	0.74	0.75
QuarterlyResults	0.82	0.53	0.64
SalesVolume	0.88	**0.75**	**0.81**
ShareRepurchase	**1.00**	**0.50**	**0.67**
TargetPrice	**1.00**	**0.75**	**0.86**
Turnover	**0.91**	**0.77**	**0.83**
avg	**0.80**	**0.71**	**0.73**

Event type	Precision	Recall	F_1-score
6B corpus GloVe one-vs-rest			
BuyRating	0.88	0.95	0.91
Debt	0.50	0.50	0.50
Dividend	0.55	0.55	0.55
MergerAcquisition	0.58	0.44	0.50
Profit	0.81	0.74	0.77
QuarterlyResults	0.84	0.47	0.60
SalesVolume	0.81	0.76	0.79
ShareRepurchase	0.75	0.50	0.60
TargetPrice	1.00	1.00	1.00
Turnover	0.94	0.65	0.77
avg	0.77	0.66	0.70

ERROR ANALYSIS

- We performed qualitative error analysis on best system prediction (SVM_Linear).
- Some event types have highly indicative lexical clues for some events:
  - “Home-serve, which also reports on Friday, rose 2.8% to pound(s) 17.54 after RBS upgraded from ”hold” to ”buy”.
    ➜ BuyRating: often triggered by upgraded, hold, buy.
- Most types show strong lexical variation.
  - “This could raise doubts about Vodafone’s target of reaching 10m subscribers by the end of the current financial year.”
  - “It will increase the number of Barclays’ customers in France by 25 per cent.”
- ➜ SalesVolume: different lexical trigger item for each possible product or service (e.g., customers, viewers, drivers).
- Ambiguous lexical clues occur in several event categories.
  - E.g., “buy”: informs MergerAcquisition and Buyrating categories.
- Classifier did not pick up on some strong lexical clues:
  - Add semantic knowledge from structured resources for semantic generalization.

CONCLUSION

- 1st gold-standard event detection dataset for micro-economic domain.

- Baseline experiments on sentence-level event typing:
  Satisfactory performance with straight-forward one-vs-rest classification method:
  SVM_Linear: 73% F1-score.
  LSTM_6B-Glove: 70% F1-score.

LIMITATIONS

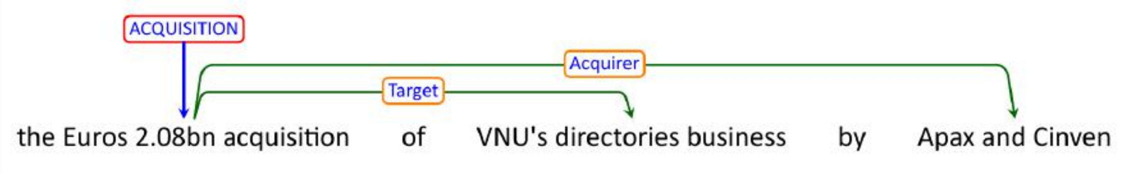
- Corpus collection:
  - Keyword-search to retrieve articles per event type: introduces lexical bias.
  - Type coverage could be improved: 18 event types in Boudhouk et al. (2016) vs. our 10 types.
- Baseline classification too elementary:
  - Task: Our data-set allows for token-span event mention detection instead of sentence-level.
  - Straight-forward classifiers: much room for improvement and more advanced approaches.

FUTURE WORK

- Collect more data: New SENTiVENT dataset for ERE-like event extraction currently being annotated.
  - Randomly crawled for 30 companies (vs. SentiFM keyword-search collection).
  - Larger type and subtype coverage.
  - Enables multi-task approach using the current SentiFM dataset.

- Task: Token-span mention detection and event subtype detection.

- Cross-lingual experiments with English and Dutch dataset.



THANK YOU! QUESTIONS?

- English SentiFM dataset and experiment replication data available here: https://osf.io/enu2k/

- Dutch + English SentiFM dataset: available upon request.

- <u>Contact:</u>

Gilles Jacobs	GillesM.Jacobs@ugent.be
Els Lefever	Els.Lefever@ugent.be
Véronique Hoste	Veronique.Hoste@ugent.be

REFERENCES

Gilles Jacobs, Els Lefever, and Véronique Hoste. 2018. Economic event detection in company-specific news text. In Proceedings of the 1st Workshop on Economics and NLP (ECONLP). ACL 2018, Melbourne, AUS, 1-10.

Craig MacKinlay. 1997. Event Studies in Economics and Finance. Journal of Economic Literature, 35(1):13–39.

Jacob Boudoukh, Ronen Feldman, Shimon Kogan, and Matthew P Richardson. 2016. Information, trading, and volatility: Evidence from firm-specific news. SSRN Working Paper 2193667.

Ronen Feldman, Benjamin Rosenfeld, Roy Bar-Haim, and Moshe Fresko. 2011. The stock sonarsentiment analysis of stocks based on a hybrid approach. In Twenty-Third IAAI Conference.

Ernest Arendarenko and Tuomo Kakkonen. 2012. Ontology-Based Information and Event Extraction for Business Intelligence. In Artificial Intelligence: Methodology, Systems, and Applications, volume 7557 of Lecture Notes in Computer Science, pages 89–102. Springer.

Alexander Hogenboom, Frederik Hogenboom, Flavius Frasincar, Kim Schouten, and Otto van der Meer. 2013. Semantics-Based Information Extraction for Detecting Economic Events. Multimedia Tools and Applications, 64(1):27–52.

Mian Du, Lidia Pivovarova, and Roman Yangarber. 2016. PULS: natural language processing for business intelligence. In Proceedings of the 2016 Workshop on Human Language Technology, pages 1–8.

Jacqueline Aguilar, Charley Beller, Paul McNamee, Benjamin Van Durme, Stephanie Strassel, Zhiyi Song, and Joe Ellis. 2014. A comparison of the events and relations across ACE, ERE, TAC-KBP, and FrameNet annotation standards. In Proceedings of the Second Workshop on EVENTS: Definition, Detection, Coreference, and Representation, pages 45–53.

Els Lefever and Veronique Hoste. 2016. A classification-based approach to economic event detection in dutch news text. In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC '16), pages 330–335. European Language Resources Association (ELRA).