

Economic event detection in company-specific news text.

Anonymous ACL submission

Abstract

This paper presents a supervised classification approach for economic event detection in English news articles. The task is conceived as a sentence-level classification task for 10 different event categories. Two different machine learning approaches were applied for the task: a rich feature set SVM approach and a word-vector-based sequence RNN-LSTM approach. We show good results for most event categories, with the linear kernel SVM outperforming the other experimental setups.

1 Introduction

In the financial domain, the way companies are perceived by investors is influenced by the news published about those companies (Engle and Ng, 1993; Tetlock, 2007; Mian and Sankaraguruswamy, 2012). Tetlock (2007), for example, tried to characterize the relationship between the content of media reports and daily stock market activity, focusing on the immediate influence of the Wall Street Journals Abreast of the Market column on U.S. stock market returns. One of his major findings was that high levels of media pessimism robustly predict downward pressure on market prices. To provide some insights into the way markets react to new information about companies, financial economists have conducted event studies.

These event studies measure the impact of a specific event on the value of a firm (MacKinlay, 1997). They offer insight into the extent to which shareholders of acquired firms gain better returns during mergers, or examine the behavior of companies stock prices around events such as dividend announcements or stock splits. Studying the impact of specific events on the stock markets, however, is a labor-intensive process, starting with the identification of a given event, the es-

timation of abnormal returns to separate the general movement of stock returns from an individual stock return, followed by a number of statistical tests seeking evidence to support the event’s economic significance. Since identifying news published about certain events in an automatic way enables researchers in the field of event studies to process more data in less time, and can consequently lead to new insights into the correlation between events and stock market movements, automatic techniques have been proposed to detect economic events in text.

Most of the existing approaches to the detection of economic events, however, are pattern-based (Arendarenko and Kakkonen, 2012; Hogenboom et al., 2013; Du et al., 2016). Rule-based ontology-driven unsupervised event detection has proven to be a popular approach in the economic domain. The Stock Sonar project (Feldman et al., 2011) notably uses domain experts to formulate event rules for ontology-driven stock sentiment analysis. This technology has been successfully used in assessing the impact of events on the stock market (Boudoukh et al., 2016) and in formulating trading strategies (Ben Ami and Feldman, 2017). Other unsupervised approaches conceptualize economic events detection as the extraction of structured event tuples (Ding et al., 2015) or as semantic frame parsing (Xie et al., 2013).

A drawback of rule-based information extraction methods is that creating rules is a difficult, time-consuming process. Furthermore, defining a set of strict rules often results in low recall scores, since these rules usually cover only a portion of the many various ways in which certain information can be lexicalized. We are not aware of any supervised classification approaches for economic event detection. However, in general domain event extraction as embodied by projects such as ACE (Ahn, 2006) and ERE/TAC-KBP (Mitamura

et al., 2016), there are plenty of supervised methods for extraction of event structures.

In this paper, we tackle the task of economic event detection by means of a machine learning approach, which we expect will be able to detect a wider variety of lexicalizations of economic events. We consider economic event detection as a sentence-level multi-class, multi-label classification task and compare two different machine learning approaches, viz. a rich feature set SVM approach, and a word-vector-based sequence RNN-LSTM approach. We show that supervised classification is a viable approach to extract economic events, with the SVM linear kernel obtaining the best classification performance.

The remainder of this paper is structured as follows. In Section 2, we present the annotated corpus of financial news articles we constructed. Section 3 introduces our two approaches to economic event detection, followed by an overview of the results in Section 4. Section 5 formulates some conclusions and ideas for future work.

2 Data description

We downloaded articles from the newspaper The Financial Times using the ProQuest Newsstand. All articles were published between November 2004 and November 2013. The articles had at least one of the following companies in the title: Barclays, BHP, Unilever, British Land, Tesco, Vodafone, BASF. All texts were pre-processed (tokenized and sentence splitted) using the LeTs Preprocess Toolkit (Van de Kauter et al., 2013). In the corpus, 10 types of company-specific economic events were manually identified, namely events regarding the *profit*, *turnover*, *sales volume*, *quarterly results*, *debt*, *target prices*, *buy ratings*, *dividend*, share repurchase, and merger/acquisition (M&A) activity of companies mentioned in the articles. Human annotators marked all mentions of each of these event types at the token level, using the brat rapid annotation tool (Stenetorp et al., 2012), a web-based tool for text annotation. The present task is sentence-level detection of event types, so one data instance can be assigned multiple event classes. Multiple labels are assigned to 3.81% ($n = 380$) of all sentence instances. An overview of the different event types and their total frequency is given in Table 1. Economic event type distribution in our datasettable.1.

Event type	# instances
No Event	7823 (75.62%)
BuyRating	227 (2.19%)
Debt	60 (0.58%)
Dividend	182 (1.76%)
MergerAcquisition	253 (2.45%)
Profit	651 (6.29%)
QuarterlyResults	267 (2.58%)
SalesVolume	487 (4.71%)
ShareRepurchase	61 (0.59%)
TargetPrice	94 (0.91%)
Turnover	240 (2.32%)
Total	9937

Table 1: Economic event type distribution in our dataset.

3 Experimental set-up

For this research, the task of economic event detection is conceived as a sentence-level multi-class classification task. We decided on comparing two different machine learning approaches: a rich feature set SVM approach, and a word-vector-based sequence RNN-LSTM approach.

The SVM approach incorporates a rich feature set with extensive semantic, syntactic, and lexical feature engineering. We built one SVM classifier per event, predicting whether the event was present in the sentence or not, in effect recasting the problem as a binary classification task. The RNN-LSTM is implemented as a multi-label single model classifier.

Performance estimation is done on a random hold-out test split (10%), whereas cross-validation experiments were carried out on the hold-in set (train set of 90%) for both hyper-parameter optimization and validation of generalization error.

Per event class, precision, recall, and F_1 -score are reported for each approach.

3.1 Support Vector Machines

For the first set of experiments, a support vector machine model was built per economic event type applying two different kernels: (1) the *Linear* kernel and (2) a hyper-parameter optimized version of the *RBF* kernel. The optimal weights for the c and g parameters for the RBF kernel were obtained by means of a 5-fold grid search on the training data for each category. All experiments were carried out with the LIBSVM package (Chang and Lin, 2011).

In a first step, the data set was linguistically pre-processed by means of the LeTs Preprocessing

Toolkit (Van de Kauter et al., 2013), which performs lemmatization, part-of-speech tagging, and named entity recognition. Consequently, a set of lexical and syntactic features were constructed on the basis of the pre-processed data.

Lexical features The following lexical features were constructed: token n-gram features (unigrams, bigrams and trigrams), character n-gram features (trigrams and fourgrams), lemma n-gram features (unigrams, bigrams and trigrams), disambiguated lemmas (lemma + associated PoS-tag), and a set of features indicating the presence of numerals, symbols, and time indicators (e.g. *yesterday*).

Syntactic features As syntactic features, we extracted three features for each PoS-category: binary (presence of category in the instance), ternary (category occurs 0, 1 or more times in the instance) and total number of occurrences of the respective PoS-label. In addition, similar features (binary, ternary, and frequency) were extracted for 6 different Named Entity types: person, organization, location, product, event, and miscellaneous.

3.2 Recurrent Neural Net LSTM

The RNN-LSTM approach was implemented using the Keras neural networks API (Chollet et al., 2015) with TensorFlow as backend (Abadi et al., 2015). We employ a straightforward neural architecture: the input-layer is a trainable embedding layer which feeds into an LSTM block. The LSTM block is connected to an output layer with sigmoid activation function. Bi-directionality of the LSTM-layer is tested in hyper-parameter optimization. We use the Adam optimization algorithm with binary cross-entropy loss function. The embedding layer turns positive integers, in our case hold-in set token indexes, in dense vectors with fixed dimensionality. An existing word embedding matrix can be used in order to use and tune pre-trained word vectors.

Three embedded inputs were tested: 200 dimensional GloVe (Pennington et al., 2014) word vectors trained on the hold-in set, 300 dimensional GloVe vectors trained on a 6 billion token corpus of Wikipedia (2014) + Gigawords5B¹ (henceforth, 6B corpus), and no pre-trained embeddings. The latter means our classifier trains embedded

¹ Available for download : <https://nlp.stanford.edu/projects/glove/>

word-representations itself based on the token sequences of the hold-in set. We evaluated our own GloVe models on an analogy quality assessment task provided with the word2vec source code. We picked the highest dimensional word vector model from the top ten ranking on the analogy task. We excluded lower dimensional vectors because preliminary tests have shown that higher dimensional pre-trained vectors obtained better scores. In further research, it is worth examining lower dimensional word vectors models as they often obtained better accuracy on the analogy quality assessment task.

The following model hyper-parameters were set by 3-fold random search with 32 iterations. The winning hyper-parameters are chosen by macro-averaged F_1 over the multi-label prediction.

RNN-LSTM hyperparameter	Setting
Bidirectionality on LSTM layer	Enabled or disabled
LSTM unit size	$d \in \{134, 268, 536\}$
Dropout rate	$r \in \{0.0, 0.2\}$
Recurrent dropout rate	$rr \in \{0.0, 0.2\}$
Batch size	$b \in \{64, 128, 256, 512\}$
Training epochs	$e \in \{32, 64, 128\}$

Table 2: RNN-LSTM model hyperparameters.

In the next section, the best model hyper-parametrization as determined by macro-averaged F_1 -score will be discussed.

4 Results and discussion

We present per class results of the SVM approach in table 3. Averaged precision, recall and F-scores per category for the optimised RBF and Linear kernel stable.3 and for the RNN-LSTM in table 4. Hold-out test Precision, recall and F-scores per category for RNN-LSTM for different word vector input table.4. Even though our classifiers were trained on a limited amount of data, we obtain good results for the detection of company-specific economic events. Overall precision scores are especially for the SVM-based approach and the RNN-LSTM with hold-in trained word vectors.

The best overall results are obtained by the linear kernel SVM which obtained far better recall than any other model. The linear kernel SVM is a comparatively fast learning algorithm and even without hyper-parameter optimization manages to obtain the best score. Including lexicon-based, syntactic, and semantic features

seems to be worthwhile when compared to the straight-forward word vector/token sequence approach used with the RNN-LSTM.

The best RNN-LSTM model is outperformed by the linear-kernel SVM approach and is on par with the optimized RBF-kernel approach. Of the three input methods tested the pre-trained GloVe vectors trained on our own dataset performed best with an F_1 -score of 0.66 on hold-out and 0.63 in 3-fold cross-validation. The GloVe vectors trained on the 6B corpus obtain worse precision but slightly better recall, resulting in a comparable F_1 -score of 0.64. Not feeding pre-trained embeddings to our network shows the worst performance of all classifiers (F_1 -score of 0.54).

Category	Precision	Recall	F_1 -score
Linear Kernel			
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
Optimized RBF			
BuyRating	0.95	0.91	0.93
Debt	0.50	1.00	0.67
Dividend	0.54	0.64	0.58
MergerAcquisition	0.00	0.00	0.00
Profit	0.80	0.76	0.78
QuarterlyResults	0.83	0.56	0.67
SalesVolume	0.94	0.65	0.77
ShareRepurchase	1.00	0.50	0.67
TargetPrice	1.00	0.75	0.86
Turnover	0.87	0.77	0.82
avg	0.74	0.65	0.67

Table 3: Averaged precision, recall and F-scores per category for the optimised RBF and Linear kernels.

5 Conclusions

This paper presents preliminary experiments for sentence-level economic event detection in English news articles. The task was approached as a supervised classification approach and two different machine learning algorithms, an SVM and RNN-LSTM learner, were applied for the task. The results show good classification performance for most event categories, with the SMV linear kernel outperforming the SVM RBF kernel and RNN-LSTM setups.

In future work, we will design a more fine-

Category	Precision	Recall	F_1 -score
Hold-in set GloVe word vectors			
BuyRating	0.91	0.91	0.91
Debt	1.00	0.50	0.67
Dividend	0.50	0.36	0.42
MergerAcquisition	0.32	0.24	0.27
Profit	0.75	0.81	0.78
QuarterlyResults	0.87	0.38	0.53
SalesVolume	0.92	0.67	0.77
ShareRepurchase	0.80	0.67	0.73
TargetPrice	1.00	0.50	0.67
Turnover	0.95	0.69	0.80
avg	0.80	0.57	0.66
6B corpus GloVe word vectors			
BuyRating	0.86	0.82	0.84
Debt	0.0	0.0	0.0
Dividend	0.50	0.55	0.52
MergerAcquisition	0.40	0.32	0.36
Profit	0.82	0.79	0.81
QuarterlyResults	0.77	0.68	0.72
SalesVolume	0.84	0.73	0.78
ShareRepurchase	1.00	0.67	0.80
TargetPrice	0.75	0.75	0.75
Turnover	0.90	0.73	0.81
avg	0.68	0.60	0.64
No pre-trained word vectors			
BuyRating	0.81	0.59	0.68
Debt	0.33	0.50	0.40
Dividend	0.75	0.55	0.63
MergerAcquisition	0.21	0.12	0.15
Profit	0.83	0.33	0.47
QuarterlyResults	0.67	0.35	0.46
SalesVolume	0.86	0.61	0.71
ShareRepurchase	0.60	0.50	0.55
TargetPrice	1.00	0.50	0.67
Turnover	0.88	0.58	0.70
avg	0.69	0.46	0.54

Table 4: Hold-out test Precision, recall and F-scores per category for RNN-LSTM for different word vector input.

grained event detection model that also extracts the span of the event below the sentence level. As feature engineering seems to pay off for the extraction of economic events, we will integrate additional linguistic information by adding semantic knowledge from structured resources such as DBpedia and dedicated ontologies for economics, as well as syntactic information extracted from dependency parses.

References

- Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon

- Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. 2015. TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from tensorflow.org. <https://www.tensorflow.org/>.
- David Ahn. 2006. The stages of event extraction. In *Proceedings of the Workshop on Annotating and Reasoning about Time and Events*. Association for Computational Linguistics, pages 1–8.
- Ernest Arendarenko and Tuomo Kakkonen. 2012. Ontology-Based Information and Event Extraction for Business Intelligence. In *Artificial Intelligence: Methodology, Systems, and Applications*. Springer, volume 7557 of *Lecture Notes in Computer Science*, pages 89–102.
- Zvi Ben Ami and Ronen Feldman. 2017. Event-based trading: Building superior trading strategies with state-of-the-art information extraction tools .
- Jacob Boudoukh, Ronen Feldman, Shimon Kogan, and Matthew P Richardson. 2016. Information, trading, and volatility: Evidence from firm-specific news .
- Chih-Chung Chang and Chih-Jen Lin. 2011. LIBSVM: A Library for Support Vector Machines. *ACM Transactions on Intelligent Systems and Technology (TIST)* 2:27:1–27:27.
- François Chollet et al. 2015. Keras. <https://github.com/fchollet/keras>.
- Xiao Ding, Yue Zhang, Ting Liu, and Junwen Duan. 2015. Deep learning for event-driven stock prediction. In *Proceedings of the 24th International Conference on Artificial Intelligence*. AAAI Press, IJCAI’15, pages 2327–2333. <http://dl.acm.org/citation.cfm?id=2832415.2832572>.
- Mian Du, Lidia Pivovarov, and Roman Yangarber. 2016. Puls: natural language processing for business intelligence. In *Proceedings of the 2016 Workshop on Human Language Technology*. Go to Print Publisher, pages 1–8.
- Robert F. Engle and Victor K. Ng. 1993. Measuring and Testing the Impact of News on Volatility. *The Journal of Finance* 48(5):1749–1778.
- 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*.
- Alexander Hogenboom, Frederik Hogenboom, Flavius Frasinca, Kim Schouten, and Otto van der Meer. 2013. Semantics-Based Information Extraction for Detecting Economic Events. *Multimedia Tools and Applications* 64(1):27–52.
- A. Craig MacKinlay. 1997. Event Studies in Economics and Finance. *Journal of Economic Literature* 35(1):13–39.
- Ghulam Mujtaba Mian and Srinivasan Sankaraguruswamy. 2012. Investor Sentiment and Stock Market Response to Earnings News. *The Accounting Review* 87(4):1357–1384.
- Teruko Mitamura, Zhengzhong Liu, and Eduard Hovy. 2016. Overview of tac-kbp 2016 event nugget track. In *Text Analysis Conference*.
- Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. Glove: Global vectors for word representation. In *Empirical Methods in Natural Language Processing (EMNLP)*. pages 1532–1543. <http://www.aclweb.org/anthology/D14-1162>.
- Pontus Stenetorp, Sampo Pyysalo, Goran Topić, Tomoko Ohta, Sophia Ananiadou, and Jun’ichi Tsujii. 2012. brat: a Web-based Tool for NLP-Assisted Text Annotation. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics (EACL’12)*. Avignon, France, pages 102–107.
- Paul C. Tetlock. 2007. Giving Content to Investor Sentiment: The Role of Media in the Stock Market. *The Journal of Finance* 62(3):1139–1168.
- Marjan Van de Kauter, Geert Coorman, Els Lefever, Bart Desmet, Lieve Macken, and Véronique Hoste. 2013. LeTs Preprocess: The multilingual LT3 linguistic preprocessing toolkit. *Computational Linguistics in the Netherlands Journal* 3:103–120.
- Boyi Xie, Rebecca J. Passonneau, Leon Wu, and Germán G. Creamer. 2013. Semantic frames to predict stock price movement. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, pages 873–883. <http://www.aclweb.org/anthology/P13-1086>.