

ALGA: Automatic Logic Gate Annotator for Building Financial News Events Detectors

Dr Julija Bainsiaksinaite
Arabesque AI
julija.bainsiaksinaite.10@ucl.ac.uk

Dr Nikolaos Kaplis
Arabesque AI
nikolaos.kaplis@arabesque.com

Prof Philip Treleaven
University College London
p.treleaven@ucl.ac.uk

Abstract

We present a new automatic data labelling framework called ALGA - Automatic Logic Gate Annotator. The framework helps to create large amounts of annotated data for training domain-specific financial news events detection classifiers quicker. ALGA framework implements a rules-based approach to annotate a training dataset. This method has following advantages: 1) unlike traditional data labelling methods, it helps to filter relevant news articles from noise; 2) allows easier transferability to other domains and better interpretability of models trained on automatically labelled data. To create this framework, we focus on the U.S.-based companies that operate in the Apparel and Footwear industry. We show that event detection classifiers trained on the data generated by our framework can achieve state-of-the-art performance in the domain-specific financial events detection task. Besides, we create a domain-specific events synonyms dictionary.

1. Introduction

A company's stock price is mostly driven by its fundamentals and release of new information. Stock market-related news might change investors' expectations and affect an asset price [1]. Financially relevant news can either be about the macro events such as economic, political, sector-specific events or company-specific events that are directly or indirectly (e.g. news about company's suppliers, competitors or customers) linked to a company. News-worthy events are shown to be associated with a company's stock price changes [2]. Therefore, an ability to automatically identify which news articles are talking about financially relevant events and measure their impact may give additional insight into a company's future financial performance.

Two main research directions analyse news impact on stock prices: event-driven and sentiment-driven.

Majority of studies in this area are focusing on building predictive models that use news sentiment [3, 4, 5] or event mentions, knowledge [2, 6, 7] to forecast stock price movements. Another research area is aimed at building models that detect financially relevant news events [8, 9]. In this work, we focus on the event detection problem and present a framework for identifying financially relevant news events, i.e. events that are likely to affect a company's stock price.

We argue that to build better event detection models, they have to be domain-specific as the importance of events across different industries varies. For example, news articles about 'successful clinical trials' affect companies operating in the Pharmaceutical sector. In contrast, news about 'store closures' would impact companies in the Retail Trade or Consumer Goods sectors. However, there is a lack of research in the domain-specific event detection area mainly due to the absence of annotated corpus for training supervised learning algorithms [8]. In the previous studies, a training corpus for supervised learning financial event detection task is usually created by either human annotators or labelling articles based on a company's stock price movements. Both methods have associated challenges. The annotators introduce human-bias, data is time-consuming to label and therefore, expensive to obtain. Meanwhile, labelling articles based on stock price movements alone is ambiguous since both price data and news articles are extremely noisy. For example, if a set of articles about a company is released during a single day and its share price changes significantly, using the traditional labelling approach adopted by most of the studies, all articles would be labelled as 'event' articles even though only a fraction of them are talking about a financially relevant event and remaining stories are irrelevant. Figure 1 demonstrates an example of such scenario. The plot shows 80 news articles published on the 4th of June 2019 that mentioned Nike (sportswear company). On the same day, Nike's stock price dipped significantly. Only 36 articles (out of 80) mentioned financially relevant event - a lawsuit

over Nike, the remaining stories were irrelevant. The challenge is how to create a labelled training dataset for the domain-specific event detection task quickly, cheaply and build event classifiers that can distinguish between the financially important news and noise.

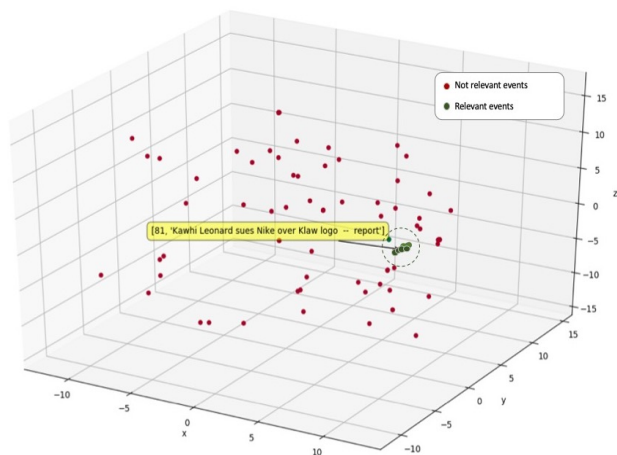


Figure 1. Example of a noisy news dataset. The plot shows 80 articles about Nike (4th of June, 2019). Only 36 stories mention the lawsuit over Nike (green) that affected the company’s stock price that day, remaining stories are irrelevant (red). After passing the articles through ALGA’s filters, the framework filters out the noise leaving relevant event articles only, then labels them accordingly. Originally, each article is represented by 768-dimensional BERT vector, for this visualisation only, we reduce the size of vectors to 3 dimensions using PCA.

The main objective of this work is to build domain-specific event classifier that can filter through noisy articles and detect stories that are likely to impact a company’s stock price. Therefore, to solve the problem mentioned above, first, we design the Automatic Logic Gate Annotator (ALGA) framework that uses a systematic approach to label training data. Second, we use this data to train an event detection classifier. We design a domain-specific event detection model for the Apparel and Footwear industry news. ALGA framework contains modular combinations of logic gates that help to filter noisy data and automatically separate articles into *event* and *non-event*. Logic gates impose conditions on an article that have to be met for it to be annotated with an *event* label (1) otherwise it is labeled with a *non-event* label (0). The conditions are combinations of the following features that are derived from articles, company’s stock prices and sector indices: 1) topic popularity, is an article talking about a popular topic on that day; 2) is a daily volume of articles higher than a threshold; 3) relevancy, is a company’s

name mentioned in the title; 4) is an article part of a cluster i.e. are there more similar articles; 5) does an article contain event synonyms from our predefined domain dictionary; 6) stock price volatility; 7) is there a significant difference between sector index and stock price returns. By passing the articles in Figure 1 through ALGA, the framework extracts only relevant events and labels them with event labels (1).

In this work, we use headlines as article representations. We embed the headlines using: BERT encoder [10], Doc2Vec [11] and TF-IDF [12]. Figure 2 shows the architecture of ALGA framework. By using this framework, we achieve state-of-the-art results in the domain-specific financial events detection task. To summarize, the contributions of this work are:

- We present ALGA - Automatic Logic Gate Annotator, an original framework for systematic news articles labelling, that can automatically generate large amounts of annotated data by separating news articles into financially relevant events and noise. Unlike studies [8, 9], we focus on English news.
- We design domain-specific news events classifier with a focus on a single financial industry.
- Using the training data generated by our framework, we train financial news events detection classifier and show state-of-the-art results.
- We compare the behaviour of different text embedding methods: Doc2Vec, TF-IDF, fine-tuned BERT (BERT encoder tuned on the financial text corpus consisting of Form 10-K reports) and general BERT (original BERT encoder). We assume that BERT embeddings should be able to generalise better for unseen words and domain-specific language (fine-tuned BERT).
- ALGA framework demonstrates scalability and transferability. As use a systematic approach to create the training dataset, the behaviour of the final financial event classifier can be easier interpretable.

This article is structured as follows. First, we review related research in Section 2. We give an overview of the data and methodology used during this study in Section 3. Sections 4 and 5 present the architecture of ALGA framework and an overview of the classifier model. We describe the results in Section 6 and conclude our work in Section 7, where we also discuss the areas for future work.

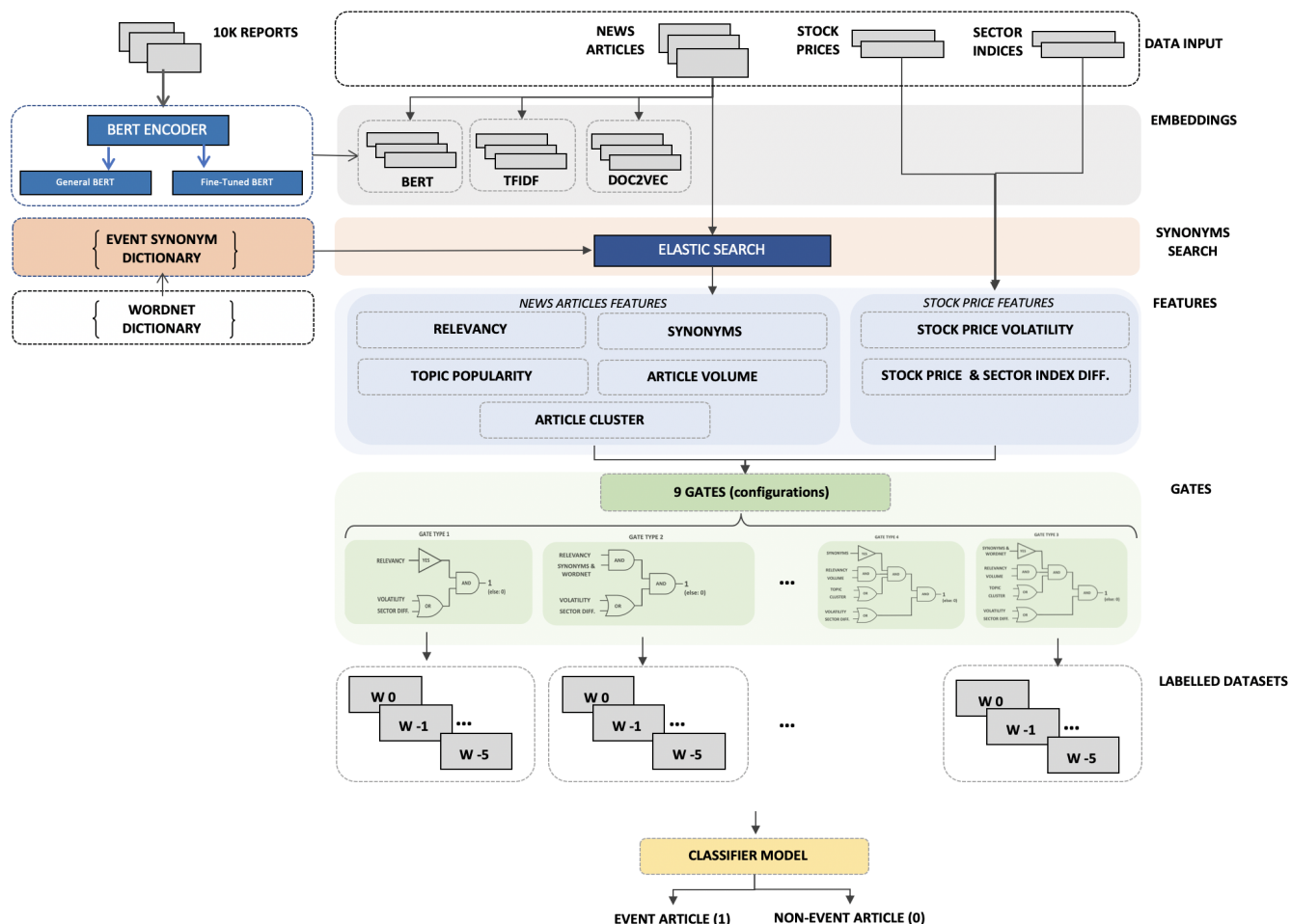


Figure 2. ALGA: Automatic Logic Gate Annotator Framework. We propose a scalable and transferable systematic method to create large amounts of labelled data for training domain-specific financial news events classifiers. The framework contains modular combinations of logic gates that help to filter noisy data and automatically separate articles into event (1) and non-event (0) news.

2. Related Work

In this section, we review the research studies that use news content to depict relationships between news and stock returns. We focus on studies that utilize news events rather than sentiment. Two main research directions exist that mainly differ by how news articles are preprocessed before they are used for further modelling: 1) articles are represented by headlines or both headlines and bodies; 2) articles are represented by event tuples $E = (Actor, Action, Object, Time)$ or article topics.

The most ubiquitous approach is to represent an article by either its title or title and body, text is

then encoded using methods such as bag-of-words [13, 14, 15], paragraph vectors [16], word (Word2Vec [17]) or sentence (Doc2Vec [18, 19, 20, 21]). Restricted Boltzmann Machines framework was used to create document vectors [22]. Recent studies built news embeddings by concatenating news titles with technical indicators [23, 24]. Research work [1] utilized a tensor-based method to integrate text and technical data instead of concatenating it. After article embeddings are created, they are used to model relationships with stock returns across different time windows, e.g. same-day closing price, next day price, etc. Researchers implemented a range of different algorithms to model this: Long Short-Term Memory

(LSTM) [16, 25, 1], bidirectional Gated Recurrent Unit (GRU) [26], Attention Mechanisms [27] together with Recurrent Neural Networks (RNN) [19, 24], hybrid Convolutions Neural Networks (CNN) models [28].

Another research direction is focusing on detecting financially relevant events in the news and uses them to model returns. One way to extract events is to use topic modelling. Research works [29, 30] analysed which topics yield non-zero stock market returns. To extract topics, scholars used Latent Dirichlet Allocation (LDA) algorithm. Another study [31] proposed density-based clustering method to extract topics.

In addition to the topic modelling, researchers followed the work [32] to design structured representation scheme for events in the form of event tuples $E = (Actor, Action, Object, Time)$ and used dependency parsing to extract events. Study [2] implemented this method to get event representations, then used Support Vector Machines (SVM) and Deep Neural Networks to model the relationship between events and stock returns. Events were encoded as sparse one-hot vectors. To reduce the sparsity of event vectors, work [6] introduced event embeddings that were trained on news corpus using Neural Tensor Network (NTN). Event embeddings can capture syntactic and semantic information but not the relationship to other similar events. Subsequent studies [7, 33, 28] improved event embeddings further by combining a knowledge graph, article bodies together with event representations to train the embeddings.

The majority of the studies mentioned above created the training data for a task by labelling the articles based on the stock price movements. The critical issue with this approach is that the news and stock price datasets are extremely noisy, and it is hard to know what indeed drives the price movements. It is essential to differentiate the relevant news from the noise and explain the reason for changes in stock prices [9]. Research work [18] tried to solve this problem by proposing an attention-based RNN model to predict returns based on the news sequence. Nevertheless, none of the studies tried to analyse the training data more in-depth to understand how well the final models can differentiate between the useful information and the noise.

Besides, there is a gap in the research for domain-specific events detection methods, especially with a focus on different financial industries mainly due to the lack of labelled news corpus for training such models. Recent study [8] developed a business events extraction framework for Chinese financial news. The method consists of the following steps: define business event dictionary (a predefined vocabulary of

business event types); expand the dictionary using word embeddings; extract news articles based on relevant event types, and using these articles find relationships between the stock prices and news events. Another study [9] designed an automatic article labelling technique for Chinese news articles and trained news events extraction model. This research work aims to fill the research gaps mentioned above.

3. Data & Methodology

In this section, we give an overview of the data and methodology used during this study. The main motivation behind our work is to build a model that classifies news articles into *financially relevant* and *noise*. We define a *financially relevant* article as a newsworthy event that is likely to affect a company's stock price. We want to identify articles that mention events such as *CEO departure*, *bribery scandal*, *factory disaster*, etc. and capture their relationship with a company's stock returns. Our objective is to build a binary classifier that takes an article as an input and outputs 1 for *financially relevant event* or 0 for *noise*. We do this by following:

1. *Automatic Logic Gate Annotator (ALGA)* - we create a rules-based approach to annotate news articles using ALGA framework which generates our training dataset. As part of the framework, we create a domain-specific event dictionary that contains synonyms associated with important events in the Apparel and Footwear industry.
2. *Domain Specific Event Classifier* - after creating the training dataset, we train Logistic Regression classifier that can distinguish between event articles from noisy irrelevant information.
3. *Results* - we evaluate the performance of the classifier and analyse the behaviour of different article embeddings.

Below we describe the datasets used during this study, article representations (embeddings), evaluation metrics together with an overview of the algorithms used for topic modelling (Latent Dirichlet Allocation) and clustering (DBSCAN).

3.1. Data

To build a domain-specific event detection model for this study, we choose 42 publicly listed U.S. companies that belong to the Apparel and Footwear industry. The datasets are collected for the period between January 2014 and September 2019 and include the following:

- Business related news articles that mention 42 companies across multiple news sources¹. In total, we obtain 488,000 articles. We use article headlines to create the following features: *relevancy*, *synonyms*, *topic popularity*, *article volume*, *article cluster*.
- Stock prices of 42 U.S. companies operating in the Apparel and Footwear industry across 2 sectors - Retail Trade and Consumer Non-Durables. We use the stock prices to create *stock price volatility* features that show how volatile a company's stock price is when an article is published.
- U.S. Retail Trade and Consumer Non-Durables sector indices. We use the sector indices in conjunction with stock prices to create *stock price and sector index difference* features that show if a company's stock price exhibits different behaviour from its sector index when an article is published.
- Form 10-K reports, in total 2,947 reports. We use the reports to fine-tune BERT encoder on the financial text corpus.

See section 4.4 for more details on how each dataset is used.

3.2. Article Representations

Research work [34] found that summarized articles predict stock price movement better than a full article, therefore for this study we choose to use a headline as a representation of an article instead of a full article. We encode article headlines using three different methods: BERT [20], Doc2Vec [17] and TF-IDF [12]. According to the study [20], words used in the financial text such as *bull* or *bear* are not associated with animals in the same way as pre-trained general language word embeddings are (c.f. [17]). Hence, it is important to introduce domain-specific language representations into a model. We utilize pre-trained language representation model BERT (Bidirectional Encoder Representations from Transformers) [10]. We fine-tune BERT model using the financial text corpus consisting of Form 10-K reports. We assume that the fine-tuned encoder should be able to generalize better for the domain-specific language, in this case finance. Two types of BERT encoders are used for further modeling - general BERT (originally trained on general language corpus) and fine-tuned BERT. We sum the last 4 layers of BERT

¹We obtain news articles for our study from EventRegistry - news aggregator service (<https://eventregistry.org>).

encoder output across all input words (i.e. an article headline) and use the result as a title embedding.

3.3. Evaluation Metrics

To evaluate the final classification models, we use the following metrics: accuracy, F1-score and ROC (Receiver Operating Characteristics) curve together with AUC (Area Under the Curve) measure. The ROC curve is a plot of the True Positive Rate (Recall) against the False Positive Rate (FPR) [35].

3.4. Topic Extraction and Clustering

To cluster articles and extract topics, we use Latent Dirichlet Allocation (LDA) [36] and Density Based Spatial Clustering of Applications with Noise (DBSCAN) [37] algorithms.

3.5. Event Dictionary

Similarly to the authors of work [20], who used Loughran McDonald's financial domain dictionary [38], we build domain-specific event dictionary with a focus on the Apparel and Footwear industry events. To build the dictionary, LDA and DBSCAN are used to generate topics and event clusters across news articles. We implement the clustering using BERT embeddings. Using the cluster analysis and extracted topics, we manually select the event-related synonyms from news articles. In total, we identify 17 event topics with a dictionary of 320 synonyms related to these topics. Figure 3 presents a snapshot of this dictionary.

4. ALGA: Automatic Logic Gate Annotator

This section describes the Automatic Logic Gate Annotator architecture in more detail and shows how we create the training data. The full framework is represented in Figure 2.

4.1. Data Input

The framework takes 3 types of data - article titles, stock prices of the 42 assets mentioned in these articles and 2 sector indices (U.S. Retail Trade and U.S. Consumer Non-durables). The titles are cleaned using standard text preprocessing techniques [29].

4.2. Embeddings

We embed 2 versions of article titles - original titles and titles with masked company name. Each title is then embedded using 4 different embedding

DEAL	MANAGEMENT
<ul style="list-style-type: none"> • Official sponsor • Licensing deal • Trade contract • Sponsorship deal • Extend partnership • Renew license 	<ul style="list-style-type: none"> • CEO departure • Appoint CCO • Get CEO • Executive turnover • Management changes • Company management
LAWSUIT	PRODUCTS
<ul style="list-style-type: none"> • Supreme Court • Court • File suit • Trademark infringement • IP breach • Fraud 	<ul style="list-style-type: none"> • Brand ambassador • Rebranding • Capsule collection • Advertising campaign • Unveil collection • Product launch

Figure 3. Domain-specific event synonyms dictionary for the Apparel and Footwear industry. As part of the study, we design a dictionary that contains 17 event topics together with 320 synonyms related to these events. The domain-specific event dictionary is subsequently used as one of the filters in ALGA framework.

methods: fine-tuned BERT, general BERT (original version), Doc2Vec and TF-IDF. In total, there are 8 variations of embedded article titles.

4.3. Synonyms Search

We upload articles to *ElasticSearch*² (optimized text search engine) which helps us to identify titles that contain words from our predefined event dictionary. We search for exact word matches as defined in the dictionary and also their synonym words by utilizing WordNet English synonyms dictionary [39]. This allows us to find articles that mention events related to 17 event categories as defined in our dictionary.

4.4. Features

Using articles, their corresponding title embeddings and financial data, the following features are created:

- *Relevancy* - we consider an article to be relevant if a company's name is mentioned in its headline and body.
- *Synonyms* - identify whether an article headline contains words from our event synonyms dictionary or their synonyms (WordNet).
- *Topic popularity* - LDA algorithm is used to extract topics from all articles. An article is

considered to contain a popular topic if it talks about a *trending* topic on that day. We define *trending* as per the following - if there are 50 articles about a company published during one day and more than 50% of them (i.e. 35) are talking about the same topic, all (35) articles are considered as *trending*.

- *Article volume* - we count the total number of articles published about a company each day. If the volume of articles during one day exceeds a predefined threshold - all articles on that day are considered as *volume* articles.
- *Article cluster* - DBSCAN is used to cluster articles in order to find similar articles. If an article falls within a cluster, it is assumed that there are more similar articles talking about a similar event/topic.
- *Stock price volatility* - we consider stock price to be volatile if standard deviations of returns for 3 or 7-day periods are higher than a predefined threshold. These periods are referred as *stock price events*.
- *Stock price and sector index difference* - we identify periods where a company's returns are significantly different (by a predefined threshold) from its sector index returns. This helps us to find the periods where a company's stock price has different behaviour from other companies within the same sector.

After creating above features for the news articles and financial data (stock prices and sector indices), both datasets are combined based on the date value. 5 different datasets are created using time shift where the financial data is shifted by the following windows: 0, -1, -2, -3 and -5 days. The intuition behind the time shift is that if an article about a *financially relevant* event is published today, can we observe an impact on a company's returns on the same day (0), after one day (-1), after two days (-2), after three days (-3) or after five days (-5).

4.5. Gates

To design an event detection model that can be applied in practice for stock trading, the model needs to be interpretable and robust to noise. By having a good understanding of data that is used to train such model, we can interpret its outputs better. For this reason, we design a rule-based systematic labelling technique using logic gates that filters and annotates training dataset.

²ElasticSearch is an optimized text search engine (<https://www.elastic.co>).

We aim to understand how different filtering rules (i.e. article volume, stock price volatility, etc.) imposed by gates impact the accuracy of the final event classification model i.e. how sensitive the final model is to training data that contains different noise levels.

We design the logic gates system that helps to filter articles based on the features defined in section 4.4. The logic gates help us to create annotated training dataset. In total, we handcraft 9 different types of logic gates configurations. Figure 4 shows an example of 5 different types of logic gates (see Appendix Figure 8 for all 9 logic gates). For example, let us consider Gate Type 2, when passing an article through this gate the article needs to meet the following conditions to be labelled as an *event* article: 1) the article has to be relevant - a company's name appears in the title and the body AND the article has to contain an event synonym from the dictionary; 2) the article has to be published during the period when the company's returns were volatile OR there was a significant difference between a company's returns and sector index returns. Both 1) AND 2) conditions have to be met for the article to be labelled as an *event* article (1), otherwise it is labelled as *non-event* (0). The same logic is applied when labelling news articles using all 9 types of gates. The output is 9 different types of training datasets. We want to analyse how different labelling conditions (gates) impact the accuracy of the final classification model trained on the data generated using the framework.

In literature, Gate Type 0 and Gate Type 1 correspond to the most commonly used methods to label training data that are based on stock price volatility and article relevance (i.e. if a company's name is mentioned in the title). The main difference between our method and the standard approach is that we label articles based on more advanced features such as sector index difference, topics, related articles (clusters) and event synonyms.

To summarize, using ALGA framework, we generate several training datasets that we use to train event classification models. The datasets contain news articles with binary labels where 1 is *financially relevant* event and 0 - *financially irrelevant* event. The logic gates help to filter noisy data. It can be observed that the more restrictive gate is, the fewer articles pass through it, hence reducing the total number of articles labelled as events. For future work, the optimal configuration of logic gates can be designed to be learnable.

5. Domain-Specific Event Classifier

Using the training datasets generated by ALGA framework, event classification models are trained. We

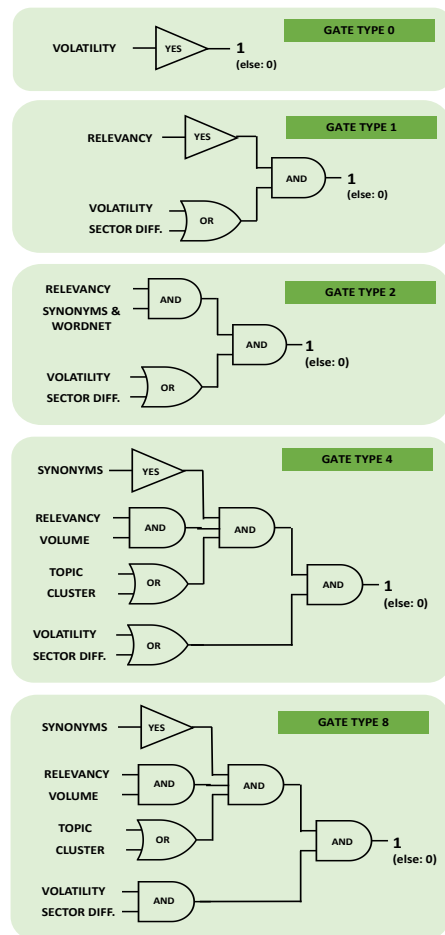


Figure 4. Logic gates systems - 9 different types of logic gates configurations are used to labels the articles. Here we show an example of 5 gates (Gate Type 0, 1, 2, 4 and 8). The logic gates are a part of ALGA framework. They are used to impose a set of conditions on an article for it to be labelled as an event (1) or non-event (0) article during the training data labelling process.

implement Logistic Regression algorithm to train the models. Considering training set \mathcal{D} with m training samples $\mathcal{D} = \{(x_i, y_i) | i = 1 : m\}$, where $x_i \in \mathbb{R}^{n+1}$ and $y \in \{0, 1\}$, Logistic Regression is defined as per equation 1 where cost function 2 is used to minimise the loss during training to find the optimal model parameters [35].

$$h_{\theta}(x_i) = \frac{1}{1 + \exp(-\theta x_i)} \quad (1)$$

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^m y_i \log(h_{\theta}(x_i)) + (1-y_i) \log(1-h_{\theta}(x_i)) \quad (2)$$

6. Results

To analyse the performance of models trained using different training datasets and title embeddings, we use the following evaluation metrics: accuracy, F-1 score, AUC-ROC curves. Besides, as a benchmark, we randomly label news articles and train classifier on the randomly annotated dataset to compare its performance against other models. Below we summarize the key results and insights from this study:

- Model trained on dataset produced by Gate Type 8 (the gate with the most restrictive conditions) has the highest accuracy. Figure 5 shows how model accuracy varies across different types of gates. Models represented in Figure 5 are trained on title embeddings with masked company names. Gate Type 0 and Gate Type 1 are the most commonly used approaches in the literature to label training data. Here, a company's news articles are labelled based on its stock price volatility (Gate Type 0) and in some cases considering company's name mention in a headline (Gate Type 1). It can be observed that the accuracy of both models that use traditional labelling approach is between 0.5 and 0.6, which is closer to the accuracy of the model trained on randomly labelled data. Meanwhile, Gate Type 4 and Gate Type 8 produce the highest accuracy - 0.85. Both gates impose the most conditions on the articles during the labelling process (see Figure 4). This shows that by implementing stricter noise filtering techniques, we can achieve better classification accuracy. F-1 scores of the models show a similar trend.
- Models trained on datasets produced by Gate Type 4, Gate Type 8 are better at separating between the two binary classes - achieving AUC above 0.88. Figure 6 shows ROC curves for models trained on fine-tuned BERT title embeddings with masked company names for window -5. Meanwhile, the model trained on data generated by Gate Type 0 achieves AUC value of approximately 0.55, which again is closer to the random behaviour of a classifier.
- Fine-tuned BERT, general BERT and TF-IDF embeddings performed similarly. Accuracies

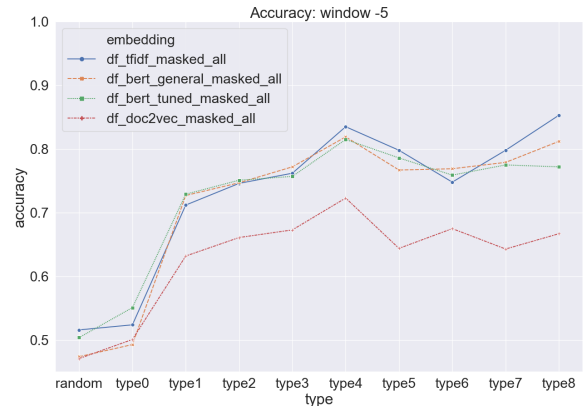


Figure 5. Accuracy of models trained on different types of datasets produced by ALGA framework. Accuracies vary across different gate types. The model trained on dataset produced by Gate Type 8, (the gate with the most conditions) has the highest accuracy. The figure shows the model accuracies for window -5 with masked companies names in the headlines.

of models trained using these embeddings are similar across all gate types, whilst the Doc2Vec accuracy is noticeably lower (see Figure 5).

- Models trained on TF-IDF title embeddings with masked company names show a better performance in differentiating between the classes, achieving AUC value of above 0.91 (see Figure 7). In comparison, both BERT embeddings (fine-tuned and general) are fluctuating around a similar AUC value of 0.9. In this experiment, models trained on TF-IDF embeddings achieved better performance than BERT embeddings. For future work, we will investigate how TF-IDF based embeddings perform when introduced to unseen words (words that were not in the original training corpus) in comparison to BERT embeddings.

7. Conclusion

In this study, we present original work in building domain-specific financial news events detection model. We introduce a new scalable and transferable automatic data labeling framework - ALGA (Automatic Logic Gate Annotator). The framework can systematically label training data for event detection classification task. For this study, ALGA framework is designed to annotate news for the Apparel and Footwear industry, although the method can be easily transferable to other

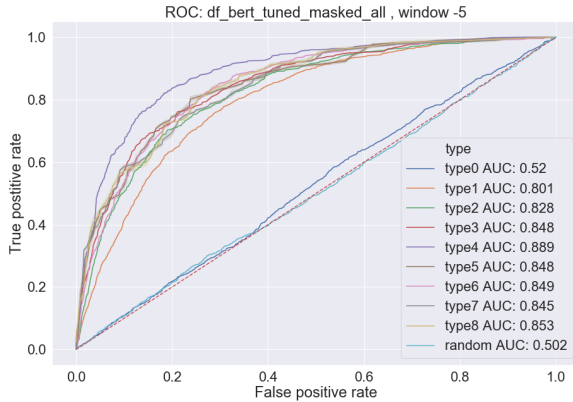


Figure 6. ROC curves for models trained on fine-tuned BERT title embeddings with masked company names and a window -5. Models trained on datasets produced by Gate Type 4, Gate Type 8 are better at separating between the two binary classes - achieving AUC above 0.88.

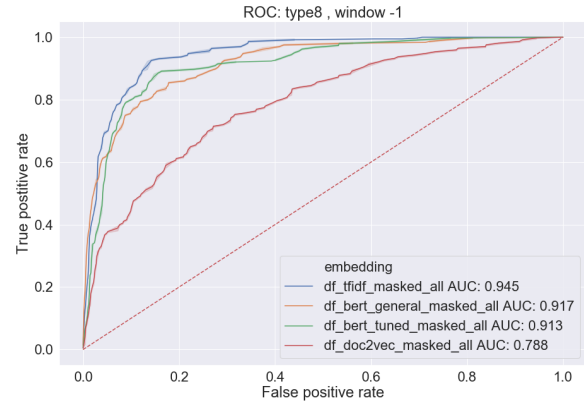


Figure 7. ROC curves for models trained on different title embeddings and data labelled by Gate Type 8 and window -1. Models trained on TF-IDF embeddings show a better performance in differentiating between the classes achieving AUC value of above 0.91. In comparison, both BERT embeddings (fine-tuned and general) are fluctuating around a similar AUC value of 0.9.

industries by using different input data and designing an industry-specific events dictionary. We show the state-of-the-art performance of our event detection model trained on data produced by ALGA framework. For future work, we will apply the framework to other industries and train different types of classification models such as Deep Neural Networks with attention layers.

References

- [1] J. Tan, J. Wang, D. Rinprasertmeechai, R. Xing, and Q. Li, "A tensor-based elstm model to predict stock price using financial news," in *Proceedings of the 52nd Hawaii International Conference on System Sciences*, 2019.
- [2] X. Ding, Y. Zhang, T. Liu, and J. Duan, "Using structured events to predict stock price movement: An empirical investigation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, (Doha, Qatar), pp. 1415–1425, Association for Computational Linguistics, Oct. 2014.
- [3] P. C. Tetlock, "Giving content to investor sentiment: The role of media in the stock market," *The Journal of Finance*, vol. 62, no. 3, pp. 1139–1168, 2007.
- [4] F. Z. Xing, E. Cambria, and R. E. Welsch, "Intelligent asset allocation via market sentiment views," *IEEE Computational Intelligence Magazine*, vol. 13, no. 4, pp. 25–34, 2018.
- [5] F. Z. Xing, E. Cambria, and Y. Zhang, "Sentiment-aware volatility forecasting," *Knowledge-Based Systems*, vol. 176, pp. 68 – 76, 2019.
- [6] X. Ding, Y. Zhang, T. Liu, and J. Duan, "Deep learning for event-driven stock prediction," in *Proceedings of the 24th International Conference on Artificial Intelligence, IJCAI'15*, pp. 2327–2333, AAAI Press, 2015.
- [7] X. Ding, Y. Zhang, T. Liu, and J. Duan, "Knowledge-driven event embedding for stock prediction," 2016.
- [8] S. Han, X. Hao, and H. Huang, "An event-extraction approach for business analysis from online chinese news," *Electronic Commerce Research and Applications*, vol. 28, pp. 244 – 260, 2018.
- [9] H. Yang, Y. Chen, K. Liu, Y. Xiao, and J. Zhao, "DCFEE: A document-level Chinese financial event extraction system based on automatically labeled training data," in *Proceedings of ACL 2018, System Demonstrations*, (Melbourne, Australia), pp. 50–55, Association for Computational Linguistics, 2018.
- [10] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: pre-training of deep bidirectional transformers for language understanding," *CoRR*, vol. abs/1810.04805, 2018.
- [11] V. Quoc and T. Mikolov, "Distributed representations of sentences and documents," *CoRR*, vol. abs/1405.4053, 2014.
- [12] K. Sprck Jones, "A statistical interpretation of term specificity and its application in retrieval," *Journal of Documentation*, vol. 28, pp. 11–21, 1972.
- [13] V. Lavrenko, M. Schmill, D. Lawrie, P. Ogilvie, D. Jensen, and J. Allan, "Mining of concurrent text and time series," in *In Proceedings of the 6th ACM SIGKDD INT'L Conference on Knowledge Discovery and Data Mining Workshop on Text Mining*, pp. 37–44, 2000.
- [14] R. P. Schumaker and H. Chen, "Textual analysis of stock market prediction using breaking financial news: The azfin text system," *ACM Trans. Inf. Syst.*, vol. 27, no. 2, pp. 12:1–12:19, 2009.
- [15] R. Luss and A. DAspremont, "Predicting abnormal returns from news using text classification," *Quantitative Finance*, vol. 15, no. 6, pp. 999–1012, 2009.

- [16] R. Akita, A. Yoshihara, T. Matsubara, and K. Uehara, "Deep learning for stock prediction using numerical and textual information," in *2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS)*, pp. 1–6, June 2016.
- [17] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *ICLR*, 2013.
- [18] Z. Hu, W. Liu, J. Bian, X. Liu, and T.-Y. Liu, "Listening to chaotic whispers: A deep learning framework for news-oriented stock trend prediction," in *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, p. 261–269, 2018.
- [19] Q. Liu, X. Cheng, S. Su, and S. Zhu, "Hierarchical complementary attention network for predicting stock price movements with news," in *Proceedings of the 27th ACM International Conference on Information and Knowledge Management, CIKM '18*, (New York, NY, USA), pp. 1603–1606, ACM, 2018.
- [20] S. Merello, A. Picasso Ratto, Y. Ma, L. Oneto, and E. Cambria, "Investigating timing and impact of news on the stock market," in *2018 IEEE International Conference on Data Mining Workshops (ICDMW)*, pp. 1348–1354, Nov 2018.
- [21] L. Yang, Z. Zhang, S. Xiong, L. Wei, J. Ng, L. Xu, and R. Dong, "Explainable text-driven neural network for stock prediction," *CoRR*, vol. abs/1902.04994, 2019.
- [22] Y. Li, T. Jin, M. Xi, S. Liu, and Z. Luo, "Massive text mining for abnormal market trend detection," *2018 IEEE International Conference on Big Data (Big Data)*, pp. 4135–4141, 2018.
- [23] P. Oncharoen and P. Vateekul, "Deep learning for stock market prediction using event embedding and technical indicators," *2018 5th International Conference on Advanced Informatics: Concept Theory and Applications (ICAICTA)*, pp. 19–24, 2018.
- [24] G. Liu and X. Wang, "A numerical-based attention method for stock market prediction with dual information," *IEEE Access*, vol. 7, pp. 7357–7367, 2019.
- [25] C.-Y. Chang, Y. Zhang, Z. Teng, Z. Bozanic, and B. Ke, "Measuring the information content of financial news," in *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, (Osaka, Japan), pp. 3216–3225, The COLING 2016 Organizing Committee, Dec. 2016.
- [26] H. D. Huynh, L. M. Dang, and D. Duong, "A new model for stock price movements prediction using deep neural network," in *Proceedings of the Eighth International Symposium on Information and Communication Technology*, p. 5762, Association for Computing Machinery, 2017.
- [27] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," *CoRR*, vol. abs/1706.03762, 2017.
- [28] S. Deng, N. Zhang, W. Zhang, J. Chen, J. Z. Pan, and H. Chen, "Knowledge-driven stock trend prediction and explanation via temporal convolutional network," in *Companion Proceedings of The 2019 World Wide Web Conference*, p. 678685, Association for Computing Machinery, 2019.
- [29] S. Feuerriegel, A. Ratku, and D. Neumann, "Analysis of how underlying topics in financial news affect stock prices using latent dirichlet allocation," in *2016 49th Hawaii International Conference on System Sciences (HICSS)*, pp. 1072–1081, Jan 2016.
- [30] S. Feuerriegel and N. Prölldochs, "Investor reaction to financial disclosures across topics: An application of latent dirichlet allocation," *CoRR*, vol. abs/1805.03308, 2018.
- [31] J. Wang, X. Wu, and L. Li, "Semantic connection based topic evolution," *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, pp. 5001–5002, 2017.
- [32] J. Kim, *Supervenience and Mind: Selected Philosophical Essays*. Cambridge University Press, 1993.
- [33] Y. Liu, Q. Zeng, H. Yang, and A. Carrio, "Stock price movement prediction from financial news with deep learning and knowledge graph embedding," in *Knowledge Management and Acquisition for Intelligent Systems* (K. Yoshida and M. Lee, eds.), (Cham), pp. 102–113, Springer International Publishing, 2018.
- [34] X. Li, H. Xie, Y. Song, S. Zhu, Q. Li, and F. L. Wang, "Does summarization help stock prediction? a news impact analysis," *IEEE Intelligent Systems*, vol. 30, pp. 26–34, May 2015.
- [35] K. Murphy, *Machine Learning. A Probabilistic Perspective*. MIT Press, 2012.
- [36] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *The Journal of Machine Learning Research*, vol. 3, pp. 993–1022, 2003.
- [37] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters a density-based algorithm for discovering clusters in large spatial databases with noise," in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, KDD'96*, pp. 226–231, AAAI Press, 1996.
- [38] T. Loughran and B. McDonald, "When is a liability not a liability? textual analysis, dictionaries, and 10-ks," *The Journal of Finance*, vol. 66, no. 1, pp. 35–65, 2011.
- [39] G. A. Miller, "Wordnet: A lexical database for english," *Commun. ACM*, vol. 38, pp. 39–41, Nov. 1995.

8. Appendix

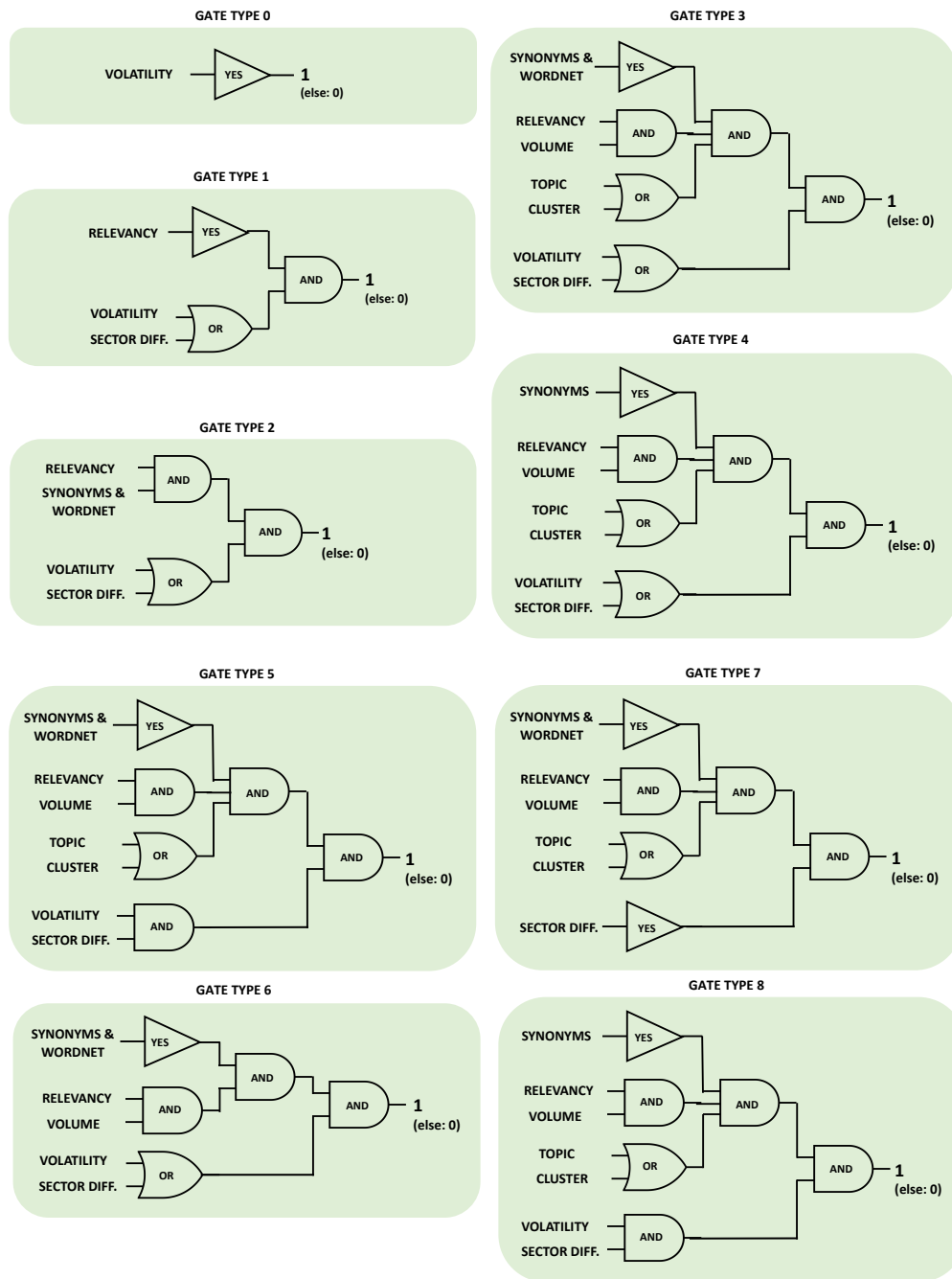


Figure 8. Logic gates systems - 9 different types of logic gates configurations that are used to label the articles. The logic gates are part of ALGA framework. They are used to impose a set of conditions on an article for it to be labelled as an event (1) or non-event (0) article during the annotation process.