Annotation model and corpus for opinion detection in economic and financial narratives

Jiahui HU*§, Patrick Paroubek†§, Dirk Schumacher*
*Natixis CIB, Paris, France
†CNRS, France
§Paris-Saclay University, France
§LISN, Bât 507, Rue du Belvedère, 91400 Orsay, France
⋆jiahui.hu@student-cs.fr
†pap@limsi.fr

Abstract

Specialized press and professional information channels influence beliefs on economic outlook or prospects for financial markets by drawing attention on particular events, and disseminating domain expert opinions. Analyzing this textual data allows for a better understanding of investors' beliefs and detecting key indicators for market dynamics.

Though considerable efforts have been made to develop data-hungry algorithms on coarse-grained level sentiment analysis on finance-related social media messages, performing fine-grained level target-dependent opinion analysis on documents written by domain experts and journalists is still a relatively unexploited field.

Since some narratives are essentially made of opinions/emotions expressed about economy and finance concepts, we address fine-grained detection of these linguistic markers at an intra-sentential level. We propose, in this paper, a global model extracting from texts terms that are specific to finance and economy or expressing an opinion/emotion in order to address the challenges of the domain-specific language we face: (1) opinions and facts about a given factor may appear at different locations (2) the range of domain-specific concepts is large and opinion may be explicit or implicit (3) syntactic structures and rhetorical relations often carry useful information for detecting market change indicators (4) emotions, like panic, also need to be detected since they are part of the economic and financial market cycle. The proposed model consists of the incorporation of fundamental approaches in natural language processing, language evaluation theory (appraisal theory), and machine learning methods for information extraction and data annotation.

In this paper, we present our annotation model and report on experiments to evaluate the quality of our dataset.

1 Introduction

The processing of information is crucial in determining financial assets' prices. Thus, market participants' opinions can be an essential driver of price dynamics. Recent progress of NLP technologies and access to digitized texts have facilitated automatic sentiment analysis of financial narratives. Most existing corpora for opinion analysis applied to economy and finance focus on the sentence level polarity (Malo et al., 2013) or text level (Cortis et al., 2017), leaving aside opinion targets. (Barbaglia et al., 2020) created a corpus focusing on the polarity of six macroeconomic aggregates, but it is not publicly available as of the time of writing. The corpus of FiQA task 1 [1] for fine-grained opinion analysis of news headlines and tweets is relatively small (1,313 samples) for training supervised learning models, and it contains relatively short sentences. In the texts we will analyze, the sentences are generally longer.

Therefore, we introduce a corpus consisting of labels annotated at the intra-sentential level by humans and algorithms to fill this gap. The novelty of our corpus is that we also consider specific rhetorical modes like financial experts do. Each sentence contains the following annotations: (1) terminologies in economics and finance, (2) opinion and emotion expressions (OEE), (3) name entities, (4) negation patterns and (5) the pair (target, polarity).

2 Methodology

2.1 Dataset

Our corpus is collected from five reputable sources from 1986 to 2021 (see Table 4 for size and sources of the raw dataset, see Figure 4 for corresponding time range). These texts aim at communicating, discussing, or commenting on business and economic activities. On the one hand, contents issued

[1] https://sites.google.com/view/fiqa/home

by the central banks [2] and corporates(MD&A of 10-K filings [3] (Ewens, 2019) and transcripts of earning calls) are first-hand information that is essential for financial markets. On the other hand, when it comes to news articles [4] and tweets, outsider comments on these official contents reflect how financial participants evaluate these events; the popularity of certain narratives in the media also shed some light on the main driver of market dynamics.

2.2 Annotation scheme

Our primary objective is to label all pairs of opinion expressions and their corresponding target, i.e. (target, opinion), at the intra-sentential level. The opinion is classified into three polarities: positive, negative and non-committal; polarities are attributed based on the judgement related to economic norms or the health of business activities. The novelty of our scheme is that we consider how financial experts communicate and analyze the evolution of event trends, namely,

a. Formulations of argumentative constructions
b. Conditional opinions
c. cause-effect relations
d. explicit speculation about the future

Our raw dataset contains both facts and opinions. In TBOA, we focus on **sentences of interest** that contain at least one terminology in economy and finance (called TOI, terms of interest) and at least one opinion & emotion expression (called OEE).

This criterion helps us to separate opinionated sentences from factual ones, because we can use existing NLP technologies to extract appraisal terms (see 2.3) and TOIs. TOIs are extracted as follows: we firstly candidate noun phrases[5] and keep just those containing elements of a domain-specific thesaurus. To extract appraisal terms, we look for exact matches between lemmatized words of each sentence and a pre-defined list of appraisal terms. The pre-annotation pipeline also includes the machine-assisted annotations of names entities and negation patterns.

2.3 How do we detect opinionated sentences

The particularity of opinionated sentences in financial narratives is that authors use evaluative language to monitor and judge an event (i.e. happenings or changes of a business or economic activity) or assess their impact. Authors may:

(1) monitor changes by using language to describe in which direction an event or a concept evolves [a],
(2) express a judgment about these dynamics by clarifying their preference; furthermore their expectations can be diversely grounded in a mix of rationality and/or emotions,
(3) and assess the intensity of these dynamics.

[a] in the DOWN & LOW category, *plummet* and *decrease* convey the notion of scaling *rapid* and *median*, respectively.

Figure 1: Our focus on specific aspects of texts written by financial experts

These elements converge toward the theoretical research about the language of evaluation. We have chosen **appraisal theory** because it provides meaning-making resources to assess the intensity (celled **Graduation**) or the direction of attitudinal expressions (called **Attitude**, i.e. affect, judgment and appreciation) and its author's commitment (called **Engagement**). As illustrated by Figure 1, we propose three axes to regroup opinion expressions about changes in economic and financial activities.

- Variation axis: gain or loss in quantity or volume, or description of stable state
- Attitude axis: recognition of value or loss in value, lack of visibility, anxious awareness of undesirable outcome; or even emotional assessment such as the intense feeling of excitement and strong desire to put ideas into practice, feelings of helplessness, the impression of losing control on the situation.
- Graduation axis is complementary to the two previous axes: high or low intensity.

2.4 Annotations

Our corpus is annotated with an open-source annotation platform called INCEpTION (Klie et al., 2018). As exemplified in Figure 2, the annotator identifies all targets towards which opinions are expressed and their polarities. Following the evaluation campaign DEFT 2018 (Paroubek et al., 2018), the annotator

(i) selects minimal information about the *Opinion & Emotion Expression* (i.e. *"dysfunctional"*, tagged [OEE]),
(ii) selects the most complete information about the target (i.e. *Sovereign bond market* in Fig-

[2] link of ECB's press conferences and speeches, link of FOMC
[3] link of MD&A data source
[4] We randomly choose sentences from Financial PhraseBank dataset (Malo et al., 2013) to apply our annotations.
[5] We use SpaCy (link), an open-source NLP toolkit for its computation efficiency.

ure 2) and attributes polarity (tagged [-] in Figure 2) to it,

(iii) then draws an unlabeled arc from [OEE] toward its corresponding target.

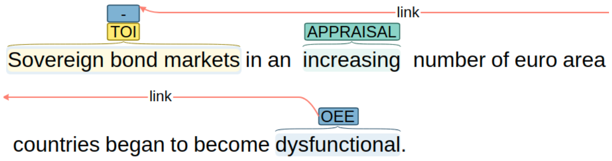

Figure 2: Example of a annotated sentence (explicit opinion)

Our corpus is annotated by one of the authors familiar with the domain terminology. Please refer to Appendix A.1 for more examples of our annotated sentences.

3 Syntactic structure of OEEs

The use of language differs from one speaker to another, depending on culture, profession, personal experience, or target audience. We assume that texts written by journalists and experts tend to use a more diverse vocabulary and syntactic structure to report facts accurately, persuade readers or polish their articles. To verify this assumption, we analyze the syntactic structure of subjective expressions in three types of texts that target different groups of people. We started from this angle because through the analysis of syntactic structure we aim to capture common phenomena in our corpus while detecting domain specificities of subjective expressions in financial narratives.

3.1 Corpora for comparison

SemEval14(Pontiki et al.) corpus is created for the NLP task[6] called *Aspect-based Sentiment Analysis*. This corpus consists of annotations of (target, polarity) of customer reviews on restaurants and laptops separately. OEEs are extracted using the neural model of (Fan et al., 2019).

MPQA(Wiebe et al., 2005) corpus consists of texts collected from a wide range of news sources. Authors annotate expressions related to opinions, beliefs, thoughts, feelings, emotions, goals, evaluations and judgments, called internal states. They divided it into two frames: expressive and direct subjectivity; the latter includes words for subjective speech events(such as say) and explicitly mentioned private states (such as fear). Our annotation scheme does not consider the language used to position a speaker's stance, corresponding to speech event expressions. Thus, we focus on the syntactic structure of expressive subjective elements, which are implicit evaluative expressions.

3.2 Tools

Numerous toolkits have been developed for syntactic analysis. We favoured Stanza(Qi et al., 2020), a state-of-the-art performance toolkit based on a neural NLP pipeline. For our algorithm, we use the universal part-of-speech (POS) tags and the syntactic dependency trees produced by the Stanza parser, focusing on the syntactic constructions of the OEEs themselves and the syntactic dependency relations that link them to other components of the sentence where they are located.

3.3 Result Analysis

We observe different patterns of opinion expressions in these four corpora.

Corpora	Unigram	2-3	4-5	6-10	>10
MPQA	28.16%	30.63%	16.61%	16.38%	8.17%
ECOFIN	49.26%	28.96%	12.11%	7.56%	1.63%
Laptop	94.26%	5.14%	0.23%	0.00%	0.00%
Restaurant	95.76%	3.24%	0.15%	0.00%	0.00%

Table 1 Statistics about number of tokens per OEE

In the SemEval 14 corpora, internauts tend to use unigrams (95.76% and 93.26% of all OEEs, respectively) for writing product reviews. Adjectives (adj) and verbs are the most used for commenting on restaurants and laptops, followed by a small portion of adverbs (adv) and nouns.

When it comes to new articles of the MPQA corpus, the variety of OEEs is the most diversified and balanced; we guess addressing implicit opinions requires more thoughtful expressions. Consequently, 72% of the OEEs are multi-grams.

In our corpus ECOFIN, the top three types of OEEs are unigrams: verbs(24.8%), adj(11.7%) and nouns(7.9%), but the overall portion of unigrams(49.26%) is much smaller than the SemEval 14. These multi-word subjective expressions, such as the combination of adj & nouns, adv & adj, are more frequently used in financial narratives than online comments (see Figure 5). In particular, the combination of verbs with other classes of words (such as adv, adposition[7]) represents at least 10% of OEEs. This observation is in line with the fact that financial experts are more likely to express their subjective opinions around changes and events, which require verbal expressions. We further investigate which are the most used verbs

[6] Semantic Evaluation 2014 Task 4

[7] preposition and postpositions

and how they relate to words in other classes(see 3.4).

Statistics of the five datasets in our experiments confirm our assumption (Table 1 and Figure 5, 6). Opinion expressions are domain-dependent; news articles and financial texts are more likely to employ multi-word opinion expressions composed of a wide range of word classes. Each word inside the OEE can modify the semantic orientation of another word, which complexifies the computation of the overall semantic orientation of the whole OEEs.

3.4 Analysis of verbs inside OEEs

Syntactic structure and word classes of OEEs can be valuable clues for determining where their corresponding target can be found. For example, for a unigram OEE whose word class is adjective, its target is likely to be the noun that follows because adjectives precede the noun they modify in English.

Following this idea, we manually examine 30 sentence of our corpus whose OEEs are in the form of verb+adp and find that most TOI precedes this type of OEEs, but some exceptions can be found in OEEs with the adp "to". Similarly, targets are very likely to be announced before OEEs "remain adj".

Recent studies ((Huang et al., 2020),(Zhao et al., 2021)) have proposed integrating syntax-related information in graph neural networks (GNN) or using GNN for sequence labelling by propagating the labelling information from known to unknown rules (which can be any rules, including syntactic ones). In the future, we want to study how these mechanism can be exploited to analyze our corpus.

3.5 Analysis of dependency relations

We also interest in the dependency relations inside each OEE of our corpus and how it is related to other words in the dependency tree. As exemplified in figure 3, the sentence is separated into three parts: OEEs, words that are above OEE (called *precedent_OEE*) and below OEE (called *posterior_OEE*).

Inside each OEE, the most frequent dependency relations are adjective and adverb modifiers and case-making relations linking adposition with the noun it attaches. The object is the fourth most important type of relation; it is connected to a verb and conveys information about the entity that undergoes a state change(see (1) in Figure 1). For example, in Figure 3, author's evaluative opinion toward *"fragmentation"* is expressed with two OEEs highlighted in orange and purple. Inside these two OEEs, the *"obj"* indicates which financial concepts (i.e., *"costs"* and *"(possibility of) economies of scale"*) are modified.

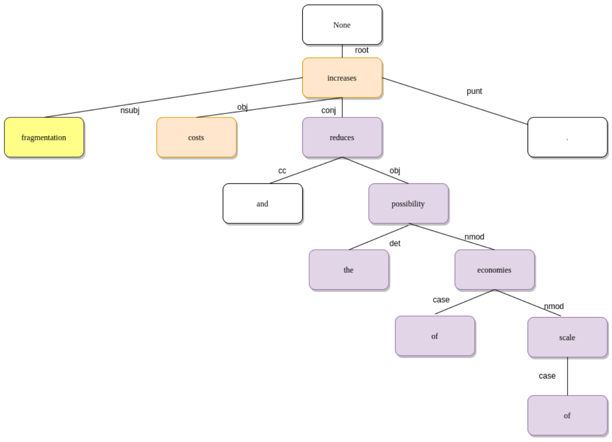

Figure 3: Dependency tree of "This fragmentation increases cost and reduces the possibility of economies of scale."

When it comes to dependency relations on the top of OEEs (posterior_OEE), we can find noun subjects, they can be the receive of an action. other most frequent dependencies of posterior_OEE and precedent_OEE can be found in Table 2.

Corpora	nsubj	obl	amod	advmod	obj	case
posterior_OEE	11.21%	8.56%	6.39%	5.52%	5.48%	3.61%
precedent_OEE	8.21%	7.53%	6.78%	3.91%	3.59%	9.0%

Table 2 Statistics about number of tokens per OEE

4 Conclusion

This paper presents our annotation scheme and the technologies used for pre-annotation. The pre-annotation output allows us to identify candidates for our corpus creation and alleviate the workload of annotators. We also compare the syntactic structure of OEEs of our financial narrative corpus with three corpora of fine-grained sentiment analysis. This comparison underlines the diversity of subjective expressions used by journalists and financial experts and the complexity of their syntactic structure. This result exemplifies why predicting TBOA from financial narratives is challenging. It also helps us understand how financial experts and journalists express opinions and how these subjective expressions in news articles and financial narratives differ from those in online comments.

In the future, we want to develop neural models adapted to our corpus by considering domain-specific knowledge, fundamental approaches in NLP in the neural model architecture to augment the machine's capacity to discover meaningful patterns.

Acknowledgements

This work is supported by the grant CIFRE, a partnership between Natixis CIB Research and the LISN Laboratory (Interdisciplinary Laboratory of Digital Sciences).

References

Luca Barbaglia, Sergio Consoli, and Sebastiano Manzan. 2020. Forecasting with economic news. *Available at SSRN 3698121*.

Keith Cortis, André Freitas, Tobias Daudert, Manuela Huerlimann, Manel Zarrouk, Siegfried Handschuh, and Brian Davis. 2017. SemEval-2017 task 5: Fine-grained sentiment analysis on financial microblogs and news. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 519–535, Vancouver, Canada. Association for Computational Linguistics.

Michael Ewens. 2019. Mda statements from public firms: 2002-2018.

Zhifang Fan, Zhen Wu, Xin-Yu Dai, Shujian Huang, and Jiajun Chen. 2019. 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)*, pages 2509–2518, Minneapolis, Minnesota. Association for Computational Linguistics.

Lianzhe Huang, Xin Sun, Sujian Li, Linhao Zhang, and Houfeng Wang. 2020. Syntax-aware graph attention network for aspect-level sentiment classification. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 799–810, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Jan-Christoph Klie, Michael Bugert, Beto Boullosa, Richard Eckart de Castilho, and Iryna Gurevych. 2018. The inception platform: Machine-assisted and knowledge-oriented interactive annotation. In *Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations*, pages 5–9. Association for Computational Linguistics.

Pekka Malo, Ankur Sinha, Pyry Takala, Pekka Korhonen, and Jyrki Wallenius. 2013. Good Debt or Bad Debt: Detecting Semantic Orientations in Economic Texts. *arXiv:1307.5336 [cs, q-fin]*. ArXiv: 1307.5336.

Patrick Paroubek, Cyril Grouin, Patrice Bellot, Vincent Claveau, Iris Eshkol-Taravella, Amel Fraisse, Agata Jackiewicz, Jihen Karoui, Laura Monceaux, and Juan-Manuel Torres-Moreno. 2018. DEFT2018 : Recherche d'information et analyse de sentiments dans des tweets concernant les transports en Île de France. In *DEFT 2018 - 14ème atelier Défi Fouille de Texte*, volume 2 of *Actes de la conférence Traitement Automatique des Langues, TALN 2018*, pages 1–11, Rennes, France.

Maria Pontiki, Dimitris Galanis, John Pavlopoulos, Harris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar. SemEval-2014 Task 4: Aspect Based Sentiment Analysis. page 9.

Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D. Manning. 2020. Stanza: A python natural language processing toolkit for many human languages.

Janyce Wiebe, Theresa Wilson, and Claire Cardie. 2005. Annotating Expressions of Opinions and Emotions in Language. *Language Resources and Evaluation*, 39(2-3):165–210.

Xinyan Zhao, Haibo Ding, and Zhe Feng. 2021. Glara: Graph-based labeling rule augmentation for weakly supervised named entity recognition. *CoRR*, abs/2104.06230.

A Appendix

Type	CB	EC	MD&A	Tweets
Percentage of sentenced of interest	26%	63%	27%	15%
Randomly chosen sentences	22,259	1,065	4,793	2,628

Table 3 Percentage of sentences of interest from randomly chosen sentences

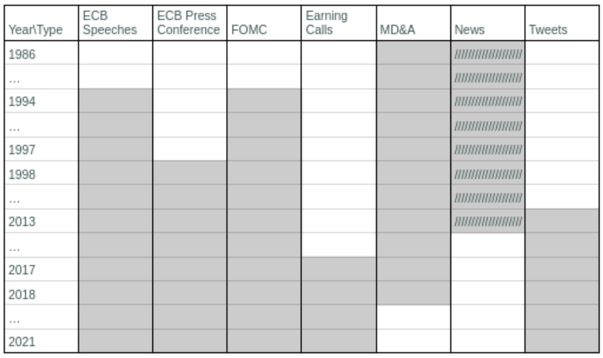

Figure 4: Shaded slashes of the column 'News' indicate that the time range of news sentences from the Financial PhraseBank dataset is incognito.

A.1 Sample sentences

Our ECOFIN corpus

(1) "The fair value of investment properties totalled EUR 2,299.9 mn , compared to EUR 2,229.5 mn in the corresponding period in 2009." [8]

(2) "This fragmentation increases costs and reduces the possibilities of economies of scale."[9]

Sent Num	target	polarity	OEE
(1)	deficit ratio	-	rise
(2)	fragmentation	-	increases costs
(2)	fragmentation	-	reduces the possibilities of economies of scale

Table 4 Our manual annotations: pair(target, polarity) and the corresponding OEE of each sample sentences

[8] from Financial Phrasebank dataset

[9] source: link

MPQA corpus[10]

- 'The criteria set by Rice are the following : the three countries in question are repressive and grave human rights violators , and aggressively seeking weapons of mass destruction.'
- 'The solidarity would bring about an international campaign to dry up the roots of terrorism and expand peace and security in the international community , but , certain countries resort to military power and embark on trampling upon human rights of civilians.'
- 'He explained that both the US and Jordan have different issues to deal with on a national level , including environmental issues.'

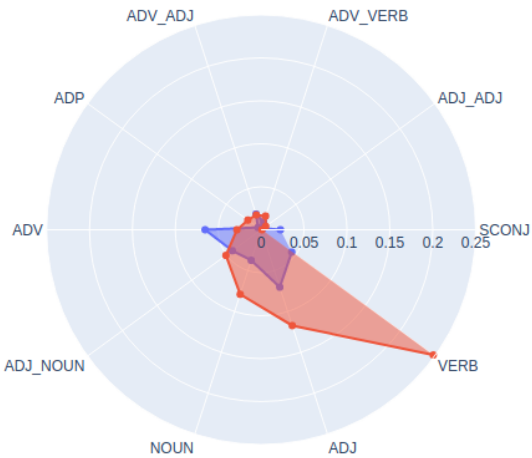

Figure 5: Top 8 universal POS of MPQA (blue) and ECOFIN (red) corpora

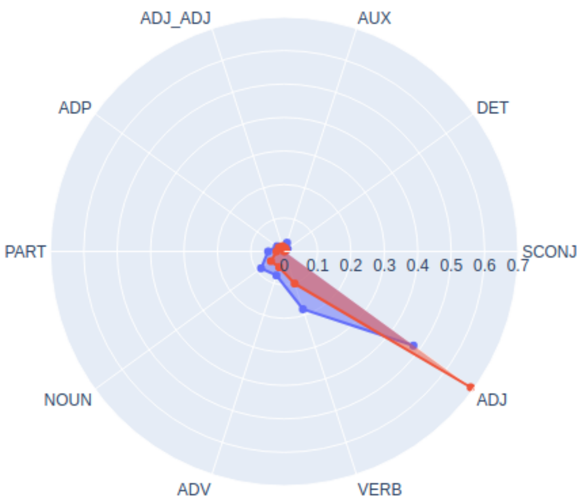

Figure 6: Top 8 universal POS of SemEval 14 corpora, Laptop (blue) and Restaurant(red)

[10]Expressive subjective elements are been underlined.