Sentiment Extraction from Financial Public Disclosure Documents

Ali Caner Türkmen

Boğaziçi University
Department of Computer Engineering
Bebek, Istanbul, Turkey 34342
caner.turkmen@boun.edu.tr

Abstract. We address the problem of extracting sentiment in financial public disclosure documents, and explore their effects on daily price movements. We take a collection of public disclosure forms submitted by four companies in the Turkish stock market. Using simple classification algorithms, we point to a significant correlation between the content of disclosure texts and the next day's price direction. We discuss the relationship between learned term weights and sentiment by comparing to a translation of a well-known financial sentiment lexicon.

1 Introduction

Using sentiment in financial news to guide investment decisions is a recent field of interest. Efficient processing of the newswire, financial commentary, social media and regulatory disclosure documents have been explored with success for forecasting price over the short and long terms.

The seminal works of Tetlock [1, 2] draw the first links between the psychosocial aspects of language in financial news and market outcomes. However, shortly thereafter, it was noted that implied sentiment of words in financial texts can differ significantly from those in generic corpora. Loughran and McDonald [3] introduced the first financial sentiment lexicon learned via statistical methods, one which was verified recently [4], and which this paper reuses.

However, sentiment lexica for financial texts are generally available in English. A method for building lexica for newly encountered corpora, contexts and languages while taking financial implications into account, is yet to be described. This is the main question addressed in this paper, where we attempt to trace the relationship between terms used in mandatory public announcements by publicly traded companies in Turkey and the market outcomes of the next day. With the end goal of building a financial sentiment vocabulary and outlining a methodology for doing so, we take a step towards processing financial news to guide accurate investment decisions.

Note that throughout this paper, we use the term "sentiment" liberally. That is, we do not necessarily point to the psychosocial aspect of terms as usually done in natural language processing, but instead focus on the statistical relationships between documents and market outcomes.

In this light, we first investigate if the presence of public disclosure filings has a statistically significant relationship with returns. We then try to associate content with meaningful financial signals using several common machine learning methods. Finally, we investigate the interpretability of learned vocabularies, and compare with a financial sentiment lexicon for English. In doing so, we explore several methods for learning both interpretable and statistically significant sentiment vocabularies, in the context of a developing stock exchange.

In the next section, we introduce the data set. In Section 3, we present the methodology and results, before concluding in Section 4.

2 Data Set

In this work, we aim to recover meaningful financial indicators from public disclosure filings by companies traded in the Turkish stock market. Public disclosure forms are mandatory announcements made through a central Public Disclosure Platform by companies traded in Istanbul's stock exchange, *Borsa Istanbul* (BIST). Among others, companies are required to report changes in ownership structure, appointment of management, disclosure of financial reports and comments on public news and rumors. In this regard, these documents are akin to Securities Exchange Commission (SEC) filings in the US.

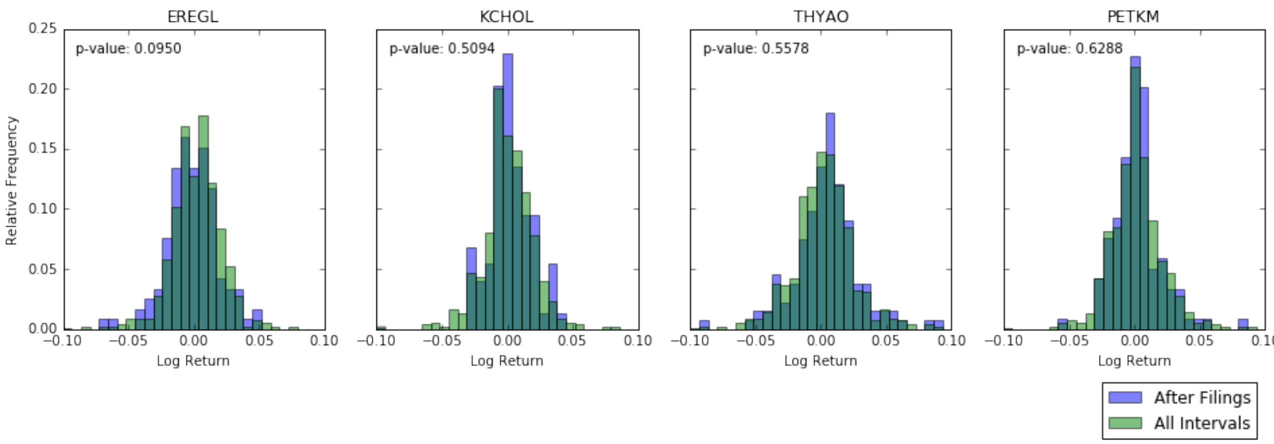
We focus exclusively on "Special Announcements" made by companies, leaving out filings such as financial reports circulated periodically. We gather announcement texts of four randomly selected companies in BIST, among constituents of the XU030 index for highest market capitalization stocks. We exclude banks since their announcements are mainly related to their market making activities. The selected stocks' ticker codes, names and industries are given in Table 1.

Table 1. Selected Stocks

Ticker Code	Company Name	Industry
EREGL	Ereğli Demir ve Çelik Fabrikalari T.A.Ş.	Metals
KCHOL	Koç Holding A.Ş.	Holding Company
THYAO	Türk Hava Yolları A.O.	Airline
PETKM	Petkim Petrokimya Holding A.Ş.	Chemicals

We take filings made during the period between November 2012 and February 2015, totaling to 551 trading days. For each company, we merge public announcements on a given day into a single document. We then label each of the documents based on the price increase/decrease of the next trading day after the announcements, with 1 for an increase, and 0 otherwise. This is due to the observation that most public announcements are filed near the closing of trading hours, and would most likely impact the next day's outcome.

Fig. 1. Histograms of log returns for the entire period and for the days after filings. Charts are annotated with p-values for two-sample t-tests. We find no significant relationship between returns and the presence of filings alone.



We exclude a widely used set of function words in the Turkish language [5], and work with a count-based term-document matrix built with a bag-of-words representation. Effectively, we formulate the question as a binary document classification problem often encountered in natural language processing.

3 Experiments

3.1 Sentiment Extraction

Before moving on to the classification problem, it is an interesting exercise to investigate if merely the appearance of a filing and the next day's price correlate. We take the distribution of log returns for the entire period, as well as those of days after filings. For each stock, we provide histograms and p-values for a two-sample t-test in Figure 1. We observe that the appearance of an announcement is not consistently followed by higher or lower returns for any of the stocks in question.

For each stock, we utilize three common "shallow" machine learning algorithms to solve the classification problem, making use of scikit-learn [6]. Finally, we combine all documents and labels in the data to investigate the existence of a common lexicon that is indicative of price movements, independent of individual stocks.

For brevity, the details of the models used will not be discussed in detail. However, we provide some details of implementation below:

– For **Logistic Regression**, we use L2 regularization, and set the penalty coefficient to 1. That is, the optimization objective is left as a simple sum of cross-entropy loss and the 2-norm of the parameter vector.

Table 2. Accuracy and Precision. 10-fold cross-validation means are presented. For each stock, the best performance is given in bold.

		EREGL	KCHOL	PETKM	THYAO	ALL
	Buy and Hold	0.5	0.48	0.52	0.46	0.49
	Buy on News	0.44	0.48	0.49	0.57	0.50
Accuracy	**Logistic Regression**	0.56	**0.63**	**0.57**	0.56	0.50
	Multinomial NB	0.58	0.60	0.51	0.58	0.49
	SVM	**0.60**	0.62	0.56	**0.60**	**0.51**
Precision	**Logistic Regression**	0.55	**0.61**	0.55	0.65	0.50
	Multinomial NB	**0.61**	0.58	0.39	0.61	0.52
	SVM	0.60	0.59	**0.56**	**0.66**	**0.52**

– We use the **Multinomial Naive Bayes** model with Laplace smoothing, see [7, p. 82].
– The **Support Vector Machine** (SVM) [8] is used with a linear kernel.

We operate on limited data, so we take several precautions to prevent overfitting. First, we do not perform hyperparameter optimization and leave hyperparameters at their most commonly used values as given above. We refrain from using deeper models and work only with simpler ones that are easy to interpret. Interpretability is also a key requirement in easily extracting a lexicon. Finally, we perform 10-fold cross-validation for each experimental setting.

We report results in terms of classification accuracy and precision [7, p. 182]. We choose these two metrics due to their unique interpretation in the context of stock market prediction. Assume the learned models were used to build an "expert system", or "strategy" that is triggered by the content of disclosure forms based on the model at hand. Disregarding trading costs, precision would correspond to the percentage of profitable *buys* in that specific stock. Accuracy, on the other hand, can be interpreted as the percentage of profitable positions if both sides of the trade (long and short) were allowed. We compare our results to two base cases. First, we report the precision of a strategy that would buy every day, i.e. buy and hold. We also report the precision of a trading algorithm that would buy every day after a news item was announced.

We present our results in Table 2. Our models outperform the base case for individual stocks. With all news items combined, we find that support vector machines are able to yield improved accuracy, although not to a significant degree. We can then reasonably hypothesize that the most discriminative terms are powerful in the context of their individual companies or industries, but that a generic lexicon cannot be recovered using simple models.

3.2 Interpreting Model Parameters

In this section, we explore the relationship between term weightings learned by one of our models and the sentiment associated to the term in financial contexts.

For this purpose, we first translate the negative terms lexicon given by Loughran & McDonald [3] into Turkish, which we make available online[1].

On the combined data set, we fit a Multinomial Naive Bayes model and estimate the log odds of a term appearing in a document followed by a decline in price. Having estimated $\hat{p}(w_i|c)$, where w_i denotes a single word and c the next day's outcome, we calculate

$$l(w_i) = \frac{\hat{p}(w_i|c=0)}{\hat{p}(w_i|c=1)}$$

We then match the lexicon of [3] to the terms used in disclosure texts. We find that for 57% of the 293 matched terms between the two vocabularies, the log-odds measure is greater than 0, i.e. it implies a decline in price. Although the majority of terms agree on their psychosocial aspects and market implications, this is only a weak correlation.

One possible explanation is that, almost surely, some of the semantics were lost during translation leading to added lexical ambiguity. One may also argue that the market may be "selling the fact", in that the expectation of negative news may have been priced in prior to the announcement. Combined with our previous argument that a statistically significant signal can be isolated in financial news, this leads to the conclusion that apparent negative meanings of terms do not necessarily lead to negative outcomes and that there may be other terms that are "bearish", but do not "sound" negative.

Upon inspecting vocabularies of [3] and those extracted by the simple rule above, we can observe some of the disagreement is indeed due to lexical ambiguity. However, we observe some terms appear to have a negative bearing, but a strong positive correlation to price. The inverse also exists, where the term is neutral despite having a strong negative implication. We give examples in Table 3. Note the appearance of words like "vote", or "retired".

4 Conclusion

In this work, we provide early evidence of the relationship between mandatory public disclosure documents and daily market outcomes in the Turkish stock market. We show that with "shallow" machine learning models, and within the context of a few randomly selected stocks, one can isolate a signal for the next day's trade direction. Under more detailed analysis of the lexicon learned by these models, we find that only a fraction of negative sounding financial terms are in fact followed by declines in price.

There are several next steps to follow this work. The first is to advance the unigram representation of this work to a more relevant language model, especially seeing that many financial terms in English translate to noun phrases in Turkish and vice versa. Second, we will expand the data and implement models

[1] github.com/canerturkmen/tr_finneg_lexicon
[2] as in legal text

Table 3. Examples of terms agreed and not agreed on by the negative sentiment lexicon in [3] and the Multinomial Naive Bayes term weights. Term weights are calculated according to the log odds, a term is considered negative when $l(w_i) > 0$.

NB Term Weight	**Positive**	**Negative**	**Negative**
Appears in [3]	**Yes**	**Yes**	**No**
Example Terms	fault penalty crisis stagnate dangerous	annulment dissident diminish fraudulent inquiry	payment retired article[2] vote temporary

capable of representing highly nonlinear relationships, in order to capture themes and higher level representations more predictive of market outcomes. Finally, we will focus on extracting information from such models in order to build a full financial sentiment lexicon in Turkish, and propose a methodology for doing so independently of language. Such an exercise will entail generalizing these models over a much wider set of stocks and news sources.

Acknowledgments

I thank Taylan Cemgil for his invaluable guidance, and the Central Securities Depository of Turkey (MKK) for making the data available.

References

1. Tetlock, P.C.: Giving content to investor sentiment: The role of media in the stock market. The Journal of Finance **62** (2007) 1139–1168
2. Tetlock, P.C., Saar-Tsechansky, M., Macskassy, S.: More than words: Quantifying language to measure firms' fundamentals. The Journal of Finance **63** (2008) 1437–1467
3. Loughran, T., McDonald, B.: When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. The Journal of Finance **66** (2011) 35–65
4. Heston, S.L., Sinha, N.R.: News versus Sentiment: Comparing Textual Processing Approaches for Predicting Stock Returns. Robert H. Smith School Research Paper (2014)
5. Amasyali, M.F., Davletov, F., Torayew, A., Ciftci, U.: Text2arff: Automatic feature extraction software for Turkish texts. In: Signal Processing and Communications Applications Conference (SIU), 2010 IEEE 18th, IEEE (2010) 629–632
6. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research **12** (2011) 2825–2830
7. Murphy, K.: Machine Learning A Probabilistic Perspective. The MIT Press (2012)
8. Cortes, C., Vapnik, V.: Support-vector networks. Machine Learning **20** (1995) 273–297