Automatic Prediction of Future Business Conditions

Lucia Noce, Alessandro Zamberletti, Ignazio Gallo, Gabriele Piccoli, and Joaquin Alfredo Rodriguez

University of Insubria,
Department of Theoretical and Applied Science,
Via Mazzini, 5, 21100 Varese, Italy
lucia.noce@uninsubria.it

Abstract. Predicting the future has been an aspiration of humans since the beginning of time. Today, predicting both macro- and micro-economic events is an important activity enabling better policy and the potential for profits. In this work, we present a novel method for automatically extracting forward looking-statement from a specific type of formal corporate documents called earning call transcripts. Our main objective is that of improving an analyst's ability to accurately forecast future events of economic relevance, over and above the predictive contribution of quantitative firm data that companies are required to produce. By exploiting both Natural Language Processing and Machine Learning techniques, our approach is stronger and more reliable than the ones commonly used in literature and it is able to accurately classify forward-looking statements without requiring any user interaction nor extensive tuning.

Keywords: Natural Language Processing, Information Retrieval, Forward-looking Statement, Earning Call

1 Introduction

It is always tough to make predictions, especially about the future; so goes the famous joke, first appearing as a Danish proverb and brought to mainstream fame when it was ascribed to Baseball manager Yogi Berra. However, although predictions are indeed difficult, we can gather substantial information about the future by elaborating increasingly available sources of digital data, such as web pages, newspapers, blogs, books, and the like. In areas where better forecasting can result in substantial profits, like stock trading, the monitoring and analysis of digital data streams [1–3] is gaining momentum. Recent works in accounting and finance demonstrate that both the quantitative data as well as the narrative/qualitative information enclosed in corporate filings (e.g., 10-K) and earning call transcripts can be successfully exploited to predict both short term firm-specific performance and future macroeconomics fluctuations [1, 2, 4–6].

An earning call transcript is the verbatim textual record of a conference call between the management of a public company, analysts, investors and/or the media to discuss the financial results the firm achieved during a specific past reporting period (e.g., quarter, fiscal year). Thus, through these earning calls, firms provide a substantial amount of future-related information to investors, journalists, policy makers and the public at large. Moreover, earning call transcripts are formal and well-formed documents that have few of the typical NLP challenges offered by other forms of communication (e.g., blog posts, tweets, etc.) For these reasons, earning calls, and company filings more generally, have been the subject of inquiry by a rapidly growing literature seeking to extract forward-looking statements. The goal is to help business analysts and policy makers to more accurately forecast firm's future behavior and performance [7–10].

Formally, a *forward-looking statement* is defined as a short sentence that contains information likely to have, or reft to, a direct effect in the foreseeable future (e.g., plans, predictions, forecasting, expectations and intentions). Recent research shows that, by analyzing the information included in these future-looking statements it is possible to improve the analysts' ability to accurately forecast future earnings, over and above the predictive contribution of quantitative firm data and consensus analysts forecasts [11, 12].

Given its relative novelty, the accounting and finance literature approaches extraction of forward-looking statements in a basic manner. For example, the classification of a sentence as a forward-looking statement occurs when it contains at least one term from a human trusted custom dictionary [13]. As a consequence, the Computer Science community can contribute to formulate better and analyze these important documents using the instruments and the knowledge provided by the Information Extraction and Information Retrieval literature. The goal of our work is to provide a simple yet effective method for combining Natural Language Processing and Supervised Machine Learning models in order to automatically extract future-looking statements from generic earning call documents. Even though the proposed method does not require any user interaction nor extensive tuning, it achieves excellent results for a dataset comprising human tagged earning calls from different corporations and it sensibly overcomes other similar approaches. Both the source code and our dataset of sentences extracted from recent earning calls and manually tagged by experts will be available for download [1].

In Section 2, we present related research work. In Sections 3 and 4, we explain in detail the process pipeline involved in our proposed method and we describe the results of our extensive experimental analysis. Finally, in Section 5, we discuss the results underlining possible future points.

2 Related works

The proposed task is deeply related to several areas, including: Natural Language Processing, Information Extraction, Information Retrieval, Data Mining,

[1] http://artelab.dicom.uninsubria.it/

Accounting and Finance. Earning calls have always been an important information source for market analysts. However, their automated processing is a new trend and a growing literature investigates whether conference calls can positively affect analysts' forecasts by helping in defining more precise indices [1, 2]. The outcome of those works prove that the exploitation of the information contained in conference calls can indeed increase analysts' ability to forecast earnings accurately, suggesting that companies release additional and significant business information during those periodical meetings.

Various and wide studies were conducted with the aim of extracting information from earning calls [6] and, recently, it has been proved that the most significant part of information that can be derived from those earning call documents can be obtained by analyzing the tones of small subset of sentences that mainly talk about future business plans: the so-called *forward-looking statements* [4, 5].

Many recent works extract forward-looking statements in a very basic manner, e.g. by classifying a sentence as forward-looking if it contains at least one term from a set of expert selected significant words [4, 5, 13, 14]. Since these kinds of approaches proved to be effective in detecting forward-looking statements from generic documents, we decided to evaluate the feasibility of a similar approach in which, however, the dictionary of terms manually identified by experts is replaced with an automatically learned one.

Surprisingly, if we move from the narrow task of extracting forward-looking statements from business documents to the more generic task of retrieving future related information from generic documents (a task known in literature as *future retrieval*), we observe that the techniques used to approach those two problems are much different, even though the basic goal is practically the same. It is indeed interesting to analyze whether the techniques commonly used in future retrieval can be successfully applied to the forward-looking sentence extraction task; for this reason, in the following paragraphs we will introduce some of the most important future retrieval works that can be found in literature.

Speaking of future retrieval works, the first who mentioned and approached this task was Baeza-Yates [15], in his work he presents the idea of extracting future temporal information from news and considers the time as a formal attribute that can be used for any information retrieval problem. Later, Jatowt *et al.* [16] proposed a query-dependent system that, exploiting future-related information in documents, supports users in detecting, summarizing and forming predictions of recurrent events related to a specific object. In another interesting work, Kawai *et al.* [17] considered automatic ways to extract future-related information from documents using Support Vector Machine (SVM) classifiers [18]. Similarly, Jatowt *et al.* [19] proposed a model-based clustering algorithm that effectively detects future events based on information extracted from a text corpus, and estimate their probabilities over time. Others works arose as extensions of the above: (i) Dias *et al.* [20] used six different learning paradigms to see through effects of temporal features upon clustering and classification of three future-related text classes; they obtained the overall best results using a SVM classifier, (ii) Matthews *et al.* [21] proposed *Time Explorer*, a system that al-

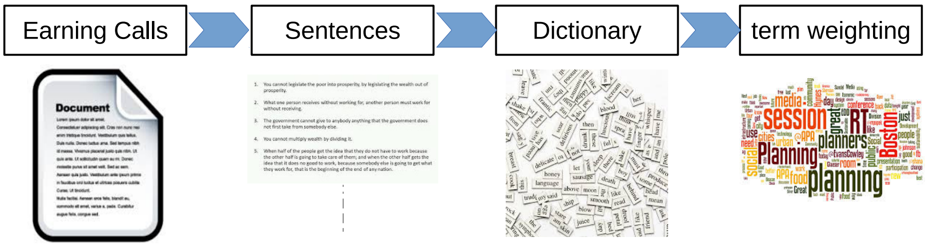

Fig. 1: A visual overview of the sequence of steps performed by the proposed model to extract forward-looking statements from earning call documents. From left to right, the starting earning call documents is split into sentences that are further split into words to build a dictionary, as in classic BOW approaches; each term in the dictionary is given a weight that is computed using Equation 1, only the terms having the highest weights appear in the final dictionary that is used to build the feature vectors provided as input to the supervised classifier model.

lows users to search information through time. It is important to point out that another slight variant of future retrieval consists in understanding the flow of people's sentiments along with temporal feature, this task is commonly called *Temporal Sentiment Analysis* [22].

Most of the above cited works adopt machine learning methods to extract future-related information from text documents, suggesting that it is indeed possible to effectively solve the problem of extracting forward-looking statements from earning call transcripts. We build on previous work but we strive to synergistically combine economic and financial studies with methods from Natural Language Processing, Information Extraction, Information Retrieval and Data Mining to provide the first automatic method of information extraction that can be used to inform economic predictions based on aggregate analysis of forward looking statements from the business community as vehicled by the earning call transcripts.

3 Proposed method

In this work, the task of identifying and extracting forward-looking statements from earning call documents has been approached as a classical text classification problem, in which each sentence is assigned one or more labels from a finite set of predefined categories. More formally, given a finite non-empty set of documents $\mathcal{X} = \{x_0, \ldots, x_n\}$ and a set of categories $\Omega = \{\omega_1, ..., \omega_c\}$, the proposed task requires to assign to every pair $(x_i, y_j) \in \mathcal{X} \times \Omega$ a boolean label yes/no. Since it is difficult and ineffective for a supervised classifier to interpret natural text as is, each document $x_i \in \mathcal{X}$ is usually represented as a vector $\phi(x_i)$ in which each element measures the number of times that a given term in x_i, contained into a finite dictionary of terms $\mathcal{D}$, appears in x_i. More in detail, given a document

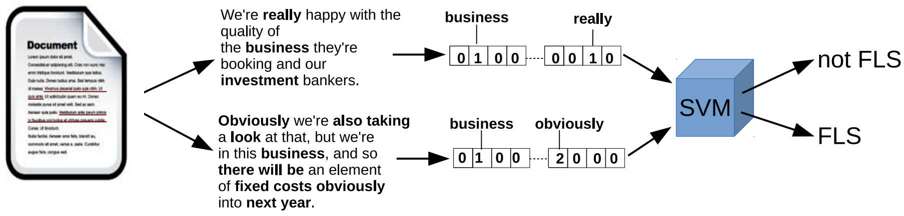

Fig. 2: An example showing how the feature vectors provided as input to the supervised classifier are generated. From left to right, the sentences extracted from the starting earning call documents are given a vectorial representation using the final weighted dictionary $\mathcal{D}$, as in Section 3; those feature vectors are finally provided as input to the SVM classifier that assigns them a label denoting wether they represent forward-looking statements or not.

$x_i \in \mathcal{X}$ and a dictionary of terms $\mathcal{D} = \{t_1, ..., t_d\}$, the document is represented as $\phi(x_i) = (x_i^1, \ldots, x_i^d)$, where each $x_i^j \in \phi(x_i)$ measures the frequency of occurrence of the term $t_j \in \mathcal{D}$ in the document $x_i \in \mathcal{X}$.

The main objective of this work is to extract forward-looking statements from earning call transcripts, for this reason, each earning call is split into a finite set of sentences that were previously tagged by different experts as either forward-looking (FLS) or not (NFLS), for this reason we have that: $\Omega = \{\text{FLS}, \text{NFLS}\}$.

As summarized in Figure 1, once the starting set of earning call documents has been split into the set of sentences $\mathcal{X}$, we proceed to further split those sentences into single terms and use them to build a primitive dictionary $\mathcal{D}'$, as in classic Bag of Words (BOW) approaches. Instead of representing each sentence as a vector using the whole dictionary $\mathcal{D}'$, we prune it to obtain the final dictionary $\mathcal{D}$ by assigning a weight to each of its element $t_i \in \mathcal{D}'$, using the weighting formula proposed by Peñas *et al.* [23], with the main goal of detecting a small subset of terms that are characteristics of forward-looking sentences. In detail, the relevance of a term $t_i \in \mathcal{D}'$ is computed as follows:

$$Relevance(t_i, sc, gc) = 1 - \frac{1}{\log_2\left(2 + \frac{F_{t_i,sc} \cdot D_{t_i,sc}}{F_{t_i,gt}}\right)} \tag{1}$$

where: (i) sc is the *specific corpus*, it corresponds to the subset of sentences from $\mathcal{X}$, extracted from the starting set of earning call documents, that were tagged by experts as FLS, (ii) gc is the *generic corpus*, it is composed by the whole set of sentences $\mathcal{X}$, (iii) $F_{t_i,sc}$ is the relative frequency of the term $t_i \in \mathcal{D}'$ in the specific corpus sc, (iv) $F_{t,gc}$ is the relative frequency of the same term $t_i \in \mathcal{D}'$ in the generic corpus gc and (v) $D_{t,sc}$ is the relative number of documents of sc in which the term $t_i \in \mathcal{D}'$ appears.

Once the relevance of every term $t_i \in \mathcal{D}'$ has been computed using Equation 1, it is possible to obtain the final smaller dictionary $\mathcal{D}$ simply by removing all those terms whose relevance value is lower than a threshold ψ, as follows:

$$\mathcal{D} = \{t_i \in \mathcal{D}' \mid Relevance(t_i, sc, gc) > \psi\} \tag{2}$$

As shown in Figure 2, given a sentence $x_i \in \mathcal{X}$ extracted from an earning call document, we compute its vectorial representation $\phi(x_i)$ as previously described, using the final dictionary $\mathcal{D}$. This vectorial representation is provided as input to a supervised machine learning algorithm with the goal of assigning one among the two possible categories in Ω to x_i. In this work, we decided to employ a Support Vector Machine (SVM) classifier [18]; this choice is motivated by the fact that the same model was used in most of the previous related works introduced in Section 2, as it can lead to optimal results with minimal tuning effort.

In summary, the proposed model is expected to take as input a sentence and to return a label that represents whether that sentence is a forward-looking statement or not. In Section 4, we present the results of an extensive experimental analysis that was conducted to identify an efficient and effective way of dealing with the forward-looking sentence extraction task. The obtained results proved that the previously described pipeline is very effective in detecting forward-looking statements; in fact, it can substantially outperform other recent methods that use hand crafted dictionaries for the same set of earning call documents without requiring any user interaction.

4 Experiments

Here we present the results obtained by performing an extensive experimental analysis of the proposed forward-looking statements identification method: (i) in Section 4.1 we provide a detailed description of the dataset used in our experiments; (ii) in Section 4.2 we present the metrics used to evaluate the effectiveness of the proposed model; (iii) in Section 4.3 we compare the results achieved by our model with those obtained by other related approaches while also providing comparisons with other techniques for building the final dictionary of terms.

4.1 Dataset

In order to build a reliable dataset of forward-looking sentences, we downloaded the earning call documents provided by three leading firms: International Business Machines Corporation (IBM), Exxon Mobil Corp (XOM) and J.P. Morgan Chase & Co. (JPM). More in detail, for each of those firms, we picked the earning call documents for the third quarter of the years 2013, 2012 and 2011.

The earning call documents were split into sentences using a classic sentence detector; those sentences were then manually labeled by a team of experts in economic and finance to identify which ones represent possible forward-looking statements. In order to determine the degree of reliability of the dataset, each sentence has been processed multiple times by different experts. By doing so, we determined that the degree of inter-rater reliability among the experts is equal

Table 1: Comparison between different parameter configurations for the classic Bag of Words (BOW) approach. The best results are obtained when considering single words ($ngram = 1$) without exploiting the position of the sentences within the earning call documents to which they belong ($context = no$).

Parameters		OA (%)	
ngram	context	prune	no-prune
1	✗	84.93	84.93
1	✓	84.57	84.57
2	✗	83.51	83.51
2	✓	81.38	81.38

to 89.73%, this high rate of accordance suggests that most of the future-looking statements in our dataset share some basic characteristics that can be easily spot by experts; this proves that, if a significant set of features is identified, the task of identifying forward-looking statements can effectively be automatized using a properly trained supervised machine learning classifier.

We collected a total amount of 3148 tagged sentences. In order to train the supervised classifier we split the dataset into train and test sets following a classic $\frac{2}{3}$-train, $\frac{1}{3}$-test split rule; this provided us with a train and test sets containing 2092 and 1046 documents respectively. It is important to point out that the sentences from the three previously cited firms were evenly split among the train and test sets in order to make the dataset as heterogeneous as possible and independent from any firms' specific language style.

4.2 Metrics

In order to compare the results of the classifier with trusted human classification, we measured the performance achieved by the different configurations of the proposed model using the Overall Accuracy (OA) metric, defined as follows:

$$\mathrm{OA} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \tag{3}$$

where (i) TP and TN are the number of true positives and true negatives, they denote the number of sentences that were correctly classified by the model as FLS or NFLS respectively, (ii) FP and FN are the number of false positives and false negatives, they denote the number of sentences that were wrongly labeled by the model in respect to their ground-truth classes.

For our overall comparison table, in addition to OA, we also measure the performance of the different algorithms using three additional metrics: *Precision* (p), *Recall* (r) and *F-measure* (f).

Table 2: Comparison between different parameter configurations for the Part of Speech (PoS) based approach. The best results can be obtained when using bigrams and defining the feature vectors using the positions in which the sentences appear within their respective earning call documents.

Parameters		OA (%)	
ngram	context	prune	no-prune
1	✗	80.50	80.50
1	✓	80.14	79.96
2	✗	80.67	80.85
2	✓	81.03	81.03

Precision represents the probability that a randomly chosen sentence $x_i \in \mathcal{X}$ gets correctly tagged by the classifier as either FLS or NFLS. *Recall* represents the probability that a randomly chosen sentence $x_i \in \mathcal{X}$ that has to be tagged as either FLS or NFLS is identified by the classifier. They are defined as follows:

$$p = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{4}$$

$$r = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{5}$$

where TP, FP and FN are defined as in Equation 3. It should be noted that *Precision* and *Recall* are not meaningful taken in isolation. In poor systems, a high p value corresponds to a small r value, and viceversa. For this reason, we also take into account the harmonic average between *Precision* and *Recall*: the *F-measure*. It is defined as follows:

$$f = \frac{2 \cdot p \cdot r}{p + r} \tag{6}$$

4.3 Results

In this section, we prove that the proposed future-looking statements identification method is able to reach and overcome other state-of-the-art approaches for the given classification task. Many experiments were conducted on the dataset described in Section 4.1; for each of them we describe the pipeline used to carry out the experiment and provide a table in which we present the obtained results while varying the different possible parameter configurations.

In our first group of experiments we used a classic Bag of Words (BOW) approach; a dictionary composed by all the words from the sentences in the training set was built. The sentences from the train and test sets were then

Table 3: Accuracies achieved by the proposed method while varying the value of the threshold parameter ψ. A significant drop in performance can be observed when moving the value of ψ from 0.3 to 0.4; the reason behind this behavior is that high values of ψ overprune the dictionary by removing terms that are highly discriminative for forward-looking statements.

threshold ψ	OA (%)
0.2	87.57
0.3	87.57
0.4	83.69
0.5	83.16

represented as BOW feature vectors and an RBF SVM [18] classifier model was trained using 5-fold cross validation. In order to find the best parameter configuration for the classifier, we ran a grid search in the following parameter space: $c \in [-5, 15]$ and $\gamma \in [-15, 3]$. In Table 1 we present the obtained results, while varying both the number of *ngram* taken into account during the building phase of the dictionary $\mathcal{D}$ and the context of a sentence into the original earning call documents. It is important to note that we did not take into account words whose length was lower than 3 characters. In detail, when using an *ngram* value of 1, we build $\mathcal{D}$ using the single words extracted from the sentences in $\mathcal{X}$. On the other hand, when using an *ngram* value of 2, we not only consider the single words but also all the possible sequences of two consecutives words inside each sentence. For example, given the following sentence: *"We plan more investements in the future"*, the first approach ($ngram = 1$) would lead to a dictionary $\mathcal{D}_{n=1}$ defined as follows: $\mathcal{D}_{n=1} = \{plan, more, investements, future\}$, while the second approach ($ngram = 2$) would generate the following dictionary of terms: $D_{n=2} = D_{n=1} \cup \{planmore, moreinvestements, investementsfuture\}$.

For each experiment, we also measure the obtained results while varying the context considered during the building phase of the feature vectors; meaning that we exploit the position of the sentences within their starting earning call documents. More specifically, given a sentence $x_i \in \mathcal{X}$ in position r within the earning call EC to which it belongs, when *context* is equal to 1, we do not take into account the information related to the position of x_i in EC while building the feature vectors. On the other hand, when *context* is equal to 2, we build the feature vector for x_i by summing the contributes from the sentences in positions $r-1$, r and $r+1$ within EC.

Moreover, for the BOW approach, we also evaluate the classification performances of the model while varying the size of the dictionary of terms $\mathcal{D}$; in detail, the dictionary is pruned by removing all its entries having frequency values lower than the median computed for all the frequencies of the terms in $\mathcal{D}$. The obtained results for this first experiment are shown in Table 1, it is possible to observe that the best OA values are achieved when ignoring the context of the

Table 4: Comparison between the results achieved by the proposed method and other approaches for the dataset described in Section 4.1, measured using the metrics introduced in Section 4.2.

method	OA (%)	p (%)	r (%)	f (%)
Bozanic *et al.* [13]	35.56	20.05	78.21	31.91
BOW	84.93	69.37	48.94	57.39
PoS	81.03	69.10	48.07	56.69
Proposed	87.57	74.82	53.23	62.21

sentences while using a dictionary built considering only unigrams. It is interesting to observe that most of the terms in the non pruned dictionary are useless, since equal results are obtained when using its pruned version. The size of the final pruned dictionary D for the best configuration of this BOW approach is equal to 2077, while the non pruned variant contains 3460 terms.

In our second experiment, we pair the previously described BOW approach with a Part of Speech (PoS) tagger algorithm to build a dictionary of tags that is used to build the feature vectors. The number of possible tags assigned to each word by the PoS tagger is equal to 42. Results are shown in Table 2, it is possible to observe that the best accuracies are achieved when building the dictionary using both 2-grams and context information. Similarly to the previous experiment, the pruning procedure does not affect the overall performances but halve the size of the final dictionary (from 2193 to 1110 terms). The results obtained by the PoS based approach are lower that the ones achieved by the previous BOW model; this suggests that the information produced by the POS tagger is not a characteristic feature of forward-looking sentences and this proves that the decisions of the experts of tagging a sentence as either FLS or NFLS are not based on the grammar of those phrases.

In the third experiment, we reproduced and tested the approach proposed by Bozanic *et al.* [13]. In order to do so, we labeled a sentence as FLS iff it contained at least one of the forward-looking terms from Table 5. Results are shown in Table 4; even though this method is quite fast if compared to the ones described in the previous experiments, it achieves poor results since the list of terms it uses is probably not discriminative enough to correctly classify sentences from documents that were never seen by the authors when they defined their dictionary of terms.

In our last experiment we evaluated the performance achieved by the proposed method, described in Section 3, while varying the value of the parameter ψ. As previously described, the latter denotes the threshold value at which the dictionary $\mathcal{D}$, weighted using the formula of Equation 1, is pruned. Results are presented in Table 3, it is possible to observe that the OA significantly drops when moving the value of ψ from 0.3 to 0.4. This is not surprising since, by taking a close look to the dictionary $\mathcal{D}$, we observed that the increment of the threshold

Table 5: Forward-looking terms from Bozanic *et al.* [13].

anticipate	continue	intends	ought
anticipated to be	continues	intent	plan
anticipates	could	intention	planning
are anticipated	estimate	intentions	plans
are anticipating	estimated to be	is anticipated	potential
are estimated	estimates	is anticipating	potentially
are estimating	expect	is estimated	predict
are expected	expectation	is estimating	predicted to be
are expecting	expected to be	is expected	predicts
are forecasted	expects	is expecting	project
are forecasting	forecast	is forecasted	projected to be
are intended	forecasted to be	is forecasting	projection
are intending	forecasts	is intended	projections
are predicted	forward	is intending	projects
are predicting	future	is predicted	schedule
are projected	goal	is predicting	scheduled
are projecting	goals	is projected	scheduled to be
are scheduled	guidance	is projecting	schedules
are scheduling	guide	is scheduled	see
are targeted	guides	is scheduling	sees
are targeting	guiding	is targeted	shall
belief	hope	is targeting	should
beliefs	hopes	may	target
believe	hoping	might	targeted to be
believes	intend	objective	targets
can	intended to be	objectives	will

value removed some terms that were highly discriminative for forward-looking statements, suggesting that the Peñas et al. [23] weighting formula can probably be improved to obtain even better results. Nevertheless, as shown in the overall comparison Table 4, the proposed method can substantially outperform the others. Using all the sentences from our training set, the proposed model requires roughly 7 min to be trained on a CPU i7 with 4gb of RAM; the final dictionary contains 1035 highly discriminative forward-looking terms. This fast training time is motivated by the fact that, in our final pipeline, we only use unigrams and we do not consider the contextual information of each sentence. Experiments using contextual information and different *ngram* values were performed, they are not reported due to lack of space; they will be made available online along with the dataset of Section 4.1. The accuracies we obtained in all those experiments were lower that the ones reported in Table 3.

5 Conclusion

The effectiveness of the proposed future-looking statements identification model has been proved by the results of our extensive experimental analysis. Even though our pipeline is very simple, it can overcome both standard methods based on human made dictionaries and other classic approaches. The choice of using the weighting function proposed by Peñas *et al.* [23] proved to be very effective in building a dictionary that well describes the most significant features of forward-looking sentences. The proposed work will be used to define more precise forecast indices to help business experts in better predicting the future fluctuations of the world economy. Moreover, the highly reliable expert tagged dataset that we have built can be used by other researchers to propose new forward-looking statements identification algorithms; this could significantly boost the interest in this narrow research field and lead to better predictions of the future business economy.

References

1. Bassemir, M., Novotny-Farkas, Z., Pachta, J.: The effect of conference calls on analysts' forecasts. European Accounting Review **22**(1) (2013) 151–183
2. Robert M. Bowen, A.K.D., Matsumoto, D.A.: Do conference calls affect analysts' forecasts? The Accounting Review **77**(2) (2002) 285–316
3. Piccoli, G., Pigni, F.: Harvesting external data: The potential of digital data streams. MIS Quarterly Executive **12**(1) (2013) 143–154
4. Li, F.: The information content of forward-looking statements in corporate filings – A Naive Bayesian Machine Learning approach. Journal of Accounting Research **48**(5) (2010) 1049–1102
5. Li, F.: The determinants and information content of the forward-looking statements in corporate filings – a naive Bayesian Machine Learning Approach. In: Proceedings of Financial Accounting and Reporting Section. (2008)
6. Muslu, V., Radhakrishnan, S., Subramanyam, K.R., Lim, D.: Forward-looking disclosures in the MD&A and the financial information environment. Management Science (2014)
7. Beyer, A., Cohen, D.A., Lys, T.Z., Walther, B.R.: The financial reporting environment: Review of the recent literature. Journal of Accounting and Economics **50**(2-3) (2010) 296–343
8. Rogers, J.L., Buskirk, A.V.: Bundled forecasts in empirical accounting research. Journal of Accounting and Economics **55**(1) (2013) 43–65
9. Ball, R., Shivakumar, L.: How much new information is there in earnings? Journal of Accounting Research **46**(5) (2008) 975–1016
10. Baginski, S. P., J.M.H., Kimbrough, M.D.: The effect of legal environment on voluntary disclosure: Evidence from management earnings forecasts issued in U.S. and Canadian markets. The Accounting Review **77**(1) (2002) 25–50
11. Frankel, R., W.M., Sun., Y.: Do pennies matter? Investor relations consequences of small negative earnings surprises. Review of Accounting Studies **15**(1) (2009) 220–242
12. Kimbrough, M.: The effect of conference calls on analyst and market underreaction to earnings announcements. The Accounting Review **80**(1) (2005) 189–219

13. Bozanic, Z., Roulstone, D.T., Van Buskirk, A.: Management earnings forecasts and forward-looking statements. (2013)
14. Wasley, C.E., Wu., J.S.: Why do managers voluntarily issue cash flow forecasts? Journal of Accounting Research **44**(2) (2006) 389–429
15. Baeza-Yates, R.: Searching the future. In: Proceedings of the ACM SIGIR workshop of Mathematical/Formal Methods in Information Retrieval. SIGIR (2005)
16. Jatowt, A., Kanazawa, K., Oyama, S., Tanaka, K.: Supporting analysis of future-related information in news archives and the web. In: Proceedings of the ACM/IEEE-CS Joint Conference on Digital Libraries. JCDL (2009)
17. Kawai, H., Jatowt, A., Tanaka, K., Kunieda, K., Yamada, K.: Chronoseeker: search engine for future and past events. In: Proceedings of the International Conference on Uniqiutous Information Management and Communication. ICUIMC (2010)
18. Cortes, C., Vapnik, V.: Support-vector networks. Machine Learning **20**(3) (1995) 273–297
19. Jatowt, A., Au Yeung, C.m.: Extracting collective expectations about the future from large text collections. In: Proceedings of the ACM International Conference on Information and Knowledge Management. CIKM (2011)
20. Dias, G., Campos, R., Jorge, A.: Future retrieval: What does the future talk about? In: Proceedings of the ACM SIGIR workshop on Enriching Information Retrieval. SIGIR (2011)
21. Matthews, M., Tolchinsky, P., Blanco, R., Atserias, J., Mika, P., Zaragoza, H.: Searching through time in the New York Times. In: Proceedings of the Symposium on Human-Computer Interaction and Information Retrieval. HCIR (2010)
22. Fukuhara, T., Nakagawa, H., Nishida, T.: Understanding sentiment of people from news articles: Temporal sentiment analysis of social events. In: Proceedings of the International Conference on Weblogs and Social Media. ICWSM (2007)
23. Penas, A., Verdejo, F., Gonzalo, J.: Corpus-based terminology extraction applied to information access. In: Proceedings of Corpus Linguistics. CL (2001)