

Sentiment analysis and the impact of employee satisfaction on firm earnings

Andy Moniz¹ and Franciska de Jong^{2,3}

¹Rotterdam School of Management, Rotterdam, The Netherlands
{moniz}@rsm.nl

²Erasmus Studio, Erasmus University, Rotterdam, The Netherlands
{fdejong}@ese.eur.nl

³Human Media Interaction, University of Twente, Enschede, The Netherlands
{f.m.g.dejong}@utwente.nl

Abstract. Prior text mining studies of corporate reputational sentiment based on newswires, blogs and Twitter feeds have mostly captured reputation from the perspective of two groups of stakeholders – the media and consumers. In this study we examine the sentiment of a potentially overlooked stakeholder group, namely, the firm’s employees. First, we present a novel dataset that uses online employee reviews to capture employee satisfaction. We employ LDA to identify salient aspects in employees’ reviews, and manually infer one latent topic that appears to be associated with the firm’s outlook. Second, we create a composite document by aggregating employee reviews for each firm and measure employee sentiment as the polarity of the composite document using the *General Inquirer* dictionary to count positive and negative terms. Finally, we define employee satisfaction as a weighted combination of the firm outlook topic cluster and employee sentiment. The results of our joint aspect-polarity model suggest that it may be beneficial for investors to incorporate a measure of employee satisfaction into their method for forecasting firm earnings.

1 Introduction

This study intends to contribute to the growing literature about applications of text mining within the field of finance. Our approach towards employees’ sentiment analysis starts from the assumption that employees are organizational assets. Management studies [1] suggest that corporate culture influences organizational behavior, especially in the areas of corporate efficiency, effectiveness and employee commitment. Indeed, according to the former CEO of IBM, “*culture is not just one aspect of the game, it is the game*” [2].

From an applications stance, our results may be of interest to investors seeking to predict firm earnings. Prior accounting research suggests that such information is not properly incorporated by the stock market due to its intangible nature, hindering the ability to measure the construct itself. To provide evidence in support of this Edmans [1] tracks the “100 Best Companies to Work for in America” published in Fortune magazine. The study posits a link between current employee satisfaction and future firm earnings that is not immediately visible to investors. We seek to complement Edmans’ work and find evidence to suggest that the forecasting power of our model is incremental to the Fortune study. We extend the regression-based approach adopted by [1] to denote the properties of an object that proxies firm outlook.

The rest of this study is structured as follows: Section 2 provides an overview of the online employee reviews dataset and highlights its advantages over the Fortune dataset. Section 3 defines employee satisfaction by developing the concepts of polarity and aspect. Throughout this paper we use the term sentiment to denote the polarity of employees’ reviews and aspect to denote the properties of an object that are commented on by reviewers. We then describe our approach to determine the classification of employee satisfaction via its impact on future firm earnings. In Section 4 we develop a polarity-only and a joint polarity-aspect model to predict firm earnings. Section 5 provides an empirical evaluation of the proposed model. We conclude in Section 6 and provide suggestions for future research.

2 The Dataset

We collected employee reviews from the career community website Glassdoor.com. The platform covers more than 250,000 global companies and contains almost 3 million anonymous salaries and reviews from 2008 onwards [3]. Reviewers provide an *Overall Score* on a scale of 1-5 and rate companies across five dimensions:

Culture & Values, Work/Life Balance, Senior Management, Comp & Benefits and Career Opportunities. Many of these ratings only begin in 2012. We extract employees' full reviews, including their perceived pros and cons of the company [4] and their 'Advice to Senior Management'. The opening sentence of reviewers' text follows a structured format, identifying whether the reviewer is a current or former employee together with the number of years' service. Comments are reviewed by website editors before publically posted. This prevents reviewers from posting defamatory attacks and from drifting off-topic that may otherwise hinder topic modelling and sentiment analysis [5] [6].

As a means to aide comparability to [1], we restrict our analysis to publically traded companies that are published in Fortune magazine's "100 Best Companies to Work for in America" list. Our corpus comprises 41,227 individual reviews, two-thirds of which were written by current employees and the remainder by former employees. The median number of reviews per company is 340, with 84% of company reviews starting in 2008.

Unlike the Fortune dataset which suffers both from untimely (annual) updates and limited data coverage, we believe that employee website comments mitigate such issues, provide a richer source of information and a novel way to look inside a company's culture [3]. Our research employs sentiment analysis using a non-proprietary dataset that we make available in open access to encourage further research¹.

3 Classification of Employee Satisfaction

The approach towards employees' sentiment analysis presented here starts from the assumption that employees are organizational assets and comprises of three steps. First, we employ Latent Dirichlet Allocation (LDA) to identify the aspects in employees' reviews and manually infer one latent topic that appears to be associated with firm outlook. Second, we measure employee sentiment as the polarity of a composite document, defined by aggregating employee reviews for each firm over each fiscal quarter. We use the *General Inquirer* dictionary to count positive and negative terms. In line with [9], our goal is not to show that a term counting method can perform as well as a Machine Learning method, but to provide a methodology to measure the impact of employee sentiment on firm earnings. Finally we define employee satisfaction as a weighted combination of firm outlook and employee sentiment. We develop a regression-based model [8][10] to forecast firm earnings by placing greater weight on documents that emphasize firm outlook.

3.1 Document

We start by defining a document as a single employee review. As the title of each document tends to summarize the review, the title and text are merged. We apply a shallow pre-processing over the text, including removal of stopwords, high frequency terms, company names and company advertisements. We use this definition of a document to train and extract the global aspects [11] of our corpus as described in Section 3.2.

We then redefine the concept of a document by combining all employee reviews written about a company into a composite document. This is because our primary goal is to evaluate the impact of aggregated employee satisfaction on firm earnings. As firms report earnings quarterly, we amalgamate² employee reviews posted during the three months' between successive quarterly earnings announcement dates. An analogous approach is adopted by [12].

3.2 Aspect

To infer salient aspects, we employ a standard implementation of LDA [13] using collapsed Gibbs sampling. Probabilistic topic models provide an unsupervised way to identify the hidden dimensions within a document and explain how much of a word in a document is related to each topic. We implement standard settings for LDA hyperparameters, $\alpha = 50/K$ and $\beta=.01$ where K is the number of topics [14]. Table 1 presents the aspects inferred by the LDA model.

¹ https://dl.dropboxusercontent.com/u/57143190/ECIR2014/employee_reviews.zip

² We require a minimum of 30 reviews [7] to form a document as a way to avoid making statistical inference on a small, potentially biased sample dataset [8].

Table 1. Topic clusters and top words identified by LDA

Representative words are the highest probability document terms for each topic cluster. The inferred aspect titles are manual annotations associated with the topic clusters.

firm outlook	development opportunities	salaries	skillset	interview tips
outlook	learn	raise	innovate	interviews
recommend	stretched	professional	individual	employers
learning	contribute	implement	specialization	private
career	ensure	costsaving	cosmetics	reviews
future	chances	solutions	skill	instructions
opportunities	career	salaries	peers	sent

Our interest lies in the first topic cluster, that we manually annotate as *firm outlook*.

3.3 Determining Sentiment

Our main resource to identify polarity is the *General Inquirer* dictionary³ [27]. The *General Inquirer* classifies words according to multiple categories, including positive and negative. This dictionary contains 1,915 positive words and 2,291 negative words. We measure polarity by counting the number of positive (P) versus negative (N) terms of a firm’s composite document [12]:

$$Polarity = (P - N)/(P + N)$$

Since former/older employees may be perversely incentivized [16] to provide negative feedback, we first statistically test for differences across different cohorts in the dataset. We compare the sentiment scores across four groups of employee reviews, distinguishing between former and current employees, junior (<5 years work experience) and senior staff (5+ years) and conduct a multivariate t-test [8] on the average sentiment scores across the four groups. We do not find a statistically significant difference in mean sentiment scores. This provides comfort that all reviews can be amalgamated into a composite document without hindering statistical inference.

3.4 Combined Approach

We adopt a statistical regression-based technique by creating a multiplicative interaction term [17] that combines *firm outlook* with sentiment. Specifically, we define the variable:

$$Outlook_sentiment_{it} = firm\ outlook_{it} \times Tone_{it}$$

The inclusion of *Outlook_sentiment* within a regression model provides a means to test that it is specifically employee sentiment related to the *firm outlook* topic cluster that is correlated to firm earnings. Our method is aligned with [18], treating positive and negative sentiment as additional topics within a LDA model.

3.5 Measuring the impact of employee satisfaction on firm earnings

Classification of employee satisfaction is challenging due to the lack of an obvious outcome to evaluate model performance [19][20][21]. The approach we take is to classify employee sentiment as positive/negative by measuring its ex-post impact on firm earnings using the concept of earnings’ surprises adopted by the financial literature [1] [10]. We first define unexpected earnings [1] for firm *i* during the financial quarter *t* as the difference between realized firm earnings (EPS_{it}) and the consensus broker estimate $E(EPS_{it})$ prior to the company’s earnings announcement. These differences are then divided by the standard deviation of broker forecasts (σ_{it}^{EPS}), so that the resulting SUE_{it} measure can be compared in the same units across all firms:

$$SUE_{it} = 1/\sigma_{it}^{EPS} \times [EPS_{it} - E(EPS_{it})]$$

³ <http://www.wjh.harvard.edu/~inquirer/homecat.htm>

The Standardized Unexpected Earnings of a firm, SUE_{it} , measures the number of standard deviations that realized earnings are above or below the consensus estimate and can be viewed as an outcome of employee satisfaction [1].

4 Model for Firm Earnings

Our primary means to evaluate the impact of employee satisfaction on firm earnings is via an ordinary least squares regression [8]. This is the standard approach adopted in financial accounting research [1] [10] [22] as a means to isolate the impact of employee satisfaction after controlling for other firm attributes. We adopt this methodology rather than more sophisticated Machine Learning techniques to aide comparability to [1]. In contrast to SVMs and neural networks, the main appeal of a regression-based approach is that the incremental forecasting power of features can readily be determined.

For a baseline, we create a naïve model that forecasts company i 's earnings surprise at time $t+1$ (the subsequent quarter) as a linear function of the company's most recent earnings surprise at time t [22]:

$$SUE_{it+1} = \beta_0 + \beta_1 SUE_{it} + \varepsilon_{it}$$

Our polarity-only model incrementally adds *Tone* to the naïve model forecast:

$$SUE_{it+1} = \beta_0 + \beta_1 SUE_{it} + \beta_2 Tone_{it} + \varepsilon_{it}$$

Finally, our joint polarity-aspect model combines both *firm outlook* and *Tone* via the multiplicative interaction term *Outlook_sentiment*. The identification of a statistically significant regression coefficient serves to test the hypothesis that a positive outlook is associated with higher than expected firm earnings over the subsequent quarter and that the feature adds incremental forecasting power to the information contained in *Tone*.

$$SUE_{it+1} = \beta_0 + \beta_1 SUE_{it} + \beta_2 Tone_{it} + \beta_3 Outlook_Sentiment_{it} + \varepsilon_{it}$$

Table 2 documents the regression results over the full sample for each model.

Table 2. Regression analysis of the models defining SUE_{it+1} as the forecast variable

<i>Model</i>	<i>Intercept</i>	<i>SUE_{it}</i>	<i>Tone_{it}</i>	<i>Outlook_Sentiment_{it}</i>
Naïve	-1.393 (-1.59)	0.230 (4.90)***		
Polarity-only	-3.338 (-2.44)	0.225 (4.79)***	4.672 (-1.85)	
Joint polarity-aspect	-3.026 (-2.23)*	0.213 (4.57)***	4.864 (-1.94)	1.435 (3.00)***

Numbers in brackets provide the test statistics. The asterisks provide the level of significance where * indicates the variable is statistically significant at the 5% level, ** at the 1% level and *** at the 0.1% level. All test statistics are based on robust standard errors [23].

Following prior financial accounting studies [24] [25], we include control variables in the regression to account for known firm attributes that may otherwise influence earnings. We include the log book-to-market ratio and the log market capitalization and the firm's prior 12 month price return. For presentation purposes only, we omit the estimated coefficients from Table 2.

The polarity-only model appears to be mildly incremental to the baseline, while the joint polarity-aspect model indicates that the interaction term is highly significant as a predictor of firm earnings.

5 Model Evaluation and Analysis

For evaluation, we select the root-mean-square error (RMSE) as a measure of the difference between the predicted model values (E_i) and the firm values actually observed (O_i):

$$RMSE = \left[\frac{1}{N} \sum_{i=1}^n (E_i - O_i)^2 \right]^{\frac{1}{2}}$$

Our choice is deemed appropriate since firm earnings are continuous rather than binary variables. We implement cross-validation using a Jack-knife approach [26] due to the limited size of our dataset (288 observations). We draw 1,000 bootstrapped samples (with replacement) using $n-1$ observations, and estimate the parameters for the regression models to predict the earnings surprise for the out-of-sample observation. The performance of the two sentiment systems are compared to the baseline. We separately identify the RMSE for positive and negative outcomes of earnings surprises.

Table 3. Comparison of RMSE across models

Model	Positive earnings surprises	Negative earnings surprises
Naïve baseline	1.823	2.952
Polarity-only	1.820	2.910
Joint polarity-aspect	1.817	2.624

The results in Table 3 show that the difference in RMSE for positive earnings surprises is negligible across the three forecast models, while RMSE for negative surprises monotonically decreases along each row and is considerably lower for joint polarity-aspect model (-11% below the Naïve baseline model). One interpretation of this result is that employee sentiment has an asymmetric effect on firm earnings. Companies with poor sentiment see negative earnings surprises during the following quarter, while companies with high employee sentiment do not see a noticeable improvement.

6 Conclusion and Future Research

To our knowledge, previous studies have only measured the impact of corporate reputation from the perception of the media and consumers. In this study, we identify a potentially neglected yet primary stakeholder of the firm and suggest that automated sentiment analysis based on employee reviews can provide a novel insight into company culture. Our findings indicate that the interaction of employee sentiment with the *firm outlook* topic cluster contains predictive power for firm earnings. This effect appears to be asymmetric, adversely affecting those companies that do not exhibit positive sentiment related to *firm outlook*.

In future work, we plan to extend our online corpus to include additional jobs and community websites and to extend coverage of companies globally. Interestingly, in an unreported principal components analysis we noticed that *firm outlook* appears to capture different dimensions to those scored by reviewers themselves. Identifying the reasons for this may be an interesting area for future classification research.

Acknowledgement

The research leading to these results has partially been supported by the Dutch national program COMMIT. The authors wish to thank Hubert Jeaneau and Julie Hudson at UBS Investment Bank for their insightful comments, and gratefully acknowledge the support of APG Asset Management.

References

- [1] Edmans, A., 2011. Does the Stock Market Fully Value Intangibles? Employee Satisfaction and Equity Prices. *Journal of Financial Economics* 101(3).
- [2] Jeaneau, H., Hudson, J., Zlotnicka, E., T., 2013. ESG Keys: Human Capital – Looking for questions.
- [3] Jeaneau, H., Hudson, J., Zlotnicka, E., T., 2013. Corporate culture: Relevant to investors? UBS Investment Research.
- [4] Kim, S. M. and Hovy, E., 2004. Determining the sentiment of opinions, in ‘Proceedings of the 20th international conference on Computational Linguistics’.
- [5] Pang, B. & Lee, L., 2004. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts, in ‘Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics’.
- [6] Hussaini, M., A. Kocyigit, D., Tapucu, B., Yanikoglu, and Y. Saygin, 2012. “An aspect-lexicon creation and evaluation tool for sentiment analysis researchers,” in ECMLPKDD.
- [7] Hogg, R., and Tanis, E., 2012. *Probability and Statistical Inference*, eighth edition.
- [8] Mardia, K.,V., Kent, J.,T., and Bibby, J.,M., 1979. *Multivariate Analysis*, Academic Press.
- [9] Pang, B., Lee, L., and Vaithyanathan, S. , 2002. Thumbs up? Sentiment classification using machine learning techniques. In *Proceedings of EMNLP-02*.
- [10] Brown, L. D., 1993. Earnings forecasting research: Its implications for capital markets research. *International Journal of Forecasting* 9: 295-320.
- [11] Titov, I. and McDonald, R. A Joint Model of Text and Aspect Ratings for Sentiment Summarization. *Proceedings of the 46th ACL*, pages 308–316, 2008.
- [12] Tetlock, P., C., 2007. Giving content to investor sentiment: The role of media in the stock market, *Journal of Finance* 62, 1139–1168.
- [13] Blei, D., M., Ng, A., Jordan, M., I., 2003. Latent Dirichlet Allocation, *Journal of Machine Learning Research* 3, 993-1022.
- [14] Griffiths, T. L., & Steyvers, M., 2004. Finding scientific topics. *Proceedings of the National Academy of Science*, 101, 5228-5235.
- [15] Kennedy, A. and D. Inkpen., 2006. Sentiment Classification of Movie Reviews using Contextual Valence Shifters. *Computational Intelligence*, vol.22(2), pp.110-125, 2006.
- [16] Tversky, A., Kahneman, D., 1973. Availability: A Heuristic for Judging Frequency and Probability. *Cognitive Psychology*, 5(2).
- [17] Brambor, T., Clark, W., R., and Golder, M., 2006. Understanding Interaction Models: Improving Empirical Analyses. *Political Analysis* 14: 63-82.
- [18] Mei, X. Shen, and C. Zhai, 2007. Automatic labelling of multinomial topic models. *SIGKDD*.
- [19] P., Turney, 2002. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews, in ‘Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics.
- [20] Wilson, T., Wiebe, J. and Hoffmann, P., 2005. Recognizing contextual polarity in phrase-level sentiment analysis, in ‘Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing’.
- [21] Ku, L. W., Lo, Y. S. and H. H., Chen., 2007. Test collection selection and gold standard generation for a multiply-annotated opinion corpus, in ‘Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions’.
- [22] Bernard, V., and Thomas, T., 1990. Evidence that stock prices do not fully reflect the implications of current earnings for future earnings. *Journal of Accounting and Economics* 13, 305-340.
- [23] White, H.,1980. A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity, *Econometrica*, 48, 817–38.
- [24] Fama, E. F., French, K. R., 1992. The cross-section of expected stock returns. *Journal of Finance* 47.
- [25] Carhart, M. M., 1997. On persistence in mutual fund performance. *Journal of Finance* 52, 57–82
- [26] Efron, B. and Tibshirani, R.J., 1993. *An Introduction to the Bootstrap*, Chapman & Hall, New York.
- [27] Stone, P., Dumphy, D. C., Smith, M. S., and Ogilvie, D. M.,1966. *The General Inquirer: A Computer Approach to Content Analysis*. The MIT Press.