

# Financial Sentiment Analysis for Risk Prediction

Chuan-Ju Wang<sup>†</sup>, Ming-Feng Tsai<sup>†</sup>, Tse Liu<sup>†</sup>, Chin-Ting Chang<sup>‡</sup>

<sup>†</sup>Department of Computer Science &  
Program in Digital Content and Technology  
National Chengchi University  
Taipei 116, Taiwan  
{mftsai, g10120}@cs.nccu.edu.tw

<sup>‡</sup>Department of Computer Science  
University of Taipei  
Taipei 100, Taiwan  
cjiang@utapei.edu.tw  
g10116006@go.utapei.edu.tw

## Abstract

This paper attempts to identify the importance of sentiment words in financial reports on financial risk. By using a finance-specific sentiment lexicon, we apply regression and ranking techniques to analyze the relations between sentiment words and financial risk. The experimental results show that, based on the bag-of-words model, models trained on sentiment words only result in comparable performance to those on origin texts, which confirms the importance of financial sentiment words on risk prediction. Furthermore, the learned models suggest strong correlations between financial sentiment words and risk of companies. As a result, these findings are of great value for providing us more insight and understanding into the impact of financial sentiment words in financial reports.

## 1 Introduction

Sentiment analysis is the task of finding the attitudes of authors about specific objects. In recent years, because of the explosion of sentiment information from social web sites (i.e., Twitter and Facebook), blogs, and online forums, sentiment analysis has become one of the popular research areas in computational linguistics, such as (Narayanan et al., 2009; Mohammad and Turney, 2010).

The growing importance of Sentiment Analysis applied to finance brings forth many research and practical issues to minds like “Why Sentiment Analysis is important?” In finance, there have been several studies (Loughran and McDonald, 2011; Price et al., 2012; Garca, 2013) using textual analysis to examine the sentiment of numerous news items, articles, financial reports, and tweets about public companies. Then, the examined sentiments can be used to reflect the correlations with other fi-

nancial measures, such as stock returns and volatilities. For most sentiment analysis algorithms, as mentioned in (Feldman, 2013), the sentiment lexicon is the most important resource. In (Loughran and McDonald, 2011), the Harvard Psychosociological Dictionary, a common dictionary for general sentiment analysis, is extended to be a finance-specific sentiment lexicon.

In this study, we attempt to use the finance-specific sentiment lexicon to model the relations between sentiment information and financial risk. In specific, we formulate the problem as two different prediction tasks: regression and ranking. For the regression task, we aim to use sentiment information to predict a company’s future risk, which is usually characterized by its real-value volatility. Instead of predicting the real-value volatility, in the ranking task, we try to employ sentiments to rank companies according to their relative risk levels. From the two tasks, we observe that, trained on the finance-specific sentiment lexicon only, both the regression models and ranking models can obtain comparable performance to those trained on original texts, even though the word dimension is largely reduced from hundreds of thousands to only one and half thousand. In addition, we also conduct some analyses on the learned models, which can provide more insight into the financial sentiments.

The remainder of this paper is organized as follows. Section 2 introduces the financial risk measure and describes the problem formulations. In Section 3, we describe the details of our experimental settings and then report the experimental results. Some discussions and analyses on the learned models are provided in Section 4. Section 5 concludes.

## 2 Methodology

### 2.1 Stock Return Volatility

In finance, volatility is a common risk metric measured by the standard deviation of a stock’s returns

over a period of time. Let  $S_t$  be the price of a stock at time  $t$ . Holding the stock for one period from time  $t - 1$  to time  $t$  would result in a simple net return:  $R_t = S_t/S_{t-1} - 1$  (Tsay, 2005). Therefore, the volatility of returns for a stock from time  $t - n$  to  $t$  can be defined as follows:

$$v_{[t-n,t]} = \sqrt{\frac{\sum_{i=t-n}^t (R_i - \bar{R})^2}{n}}, \quad (1)$$

where  $\bar{R} = \sum_{i=t-n}^t R_i / (n + 1)$ .

## 2.2 Financial Sentiment Lexicon

For most sentiment analysis algorithms, a sentiment lexicon is the most crucial resource. As mentioned in (Loughran and McDonald, 2011), a general purpose sentiment lexicon might misclassify common words in financial texts. As shown in their paper, almost three-fourths of the words in the 10-K financial reports from year 1994 to 2008, which are identified as negative by the widely used Harvard Psychosociological Dictionary, are typically not considered negative in financial contexts.

In this paper, we use a finance-specific lexicon that consists of the 6 word lists provided by (Loughran and McDonald, 2011) to analyze the relations between these sentiment words and financial risk. The six lists are shown as follows:<sup>1</sup>

1. Fin-Neg: negative business terminologies (e.g., *deficit*, *default*).
2. Fin-Pos: positive business terminologies (e.g., *achieve*, *profit*).
3. Fin-Unc: words denoting uncertainty, with emphasis on the general notion of imprecision rather than exclusively focusing on risk (e.g., *appear*, *doubt*).
4. Fin-Lit: words reflecting a propensity for legal contest or, per our label, litigiousness (e.g., *amend*, *forbear*).
5. MW-Strong (Strong Modal Words): words expressing strong levels of confidence (e.g., *always*, *must*).
6. MW-Weak (Weak Modal Words): words expressing weak levels of confidence (e.g., *could*, *might*).

<sup>1</sup>All these lists are available at [http://www.nd.edu/mcdonald/Word\\_Lists.html](http://www.nd.edu/mcdonald/Word_Lists.html).

## 2.3 Problem Formulation

### 2.3.1 Regression Task

Given a collection of financial reports  $D = \{\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_n\}$ , in which each  $\mathbf{d}_i \in \mathbb{R}^p$  and is associated with a company  $c_i$ , we seek to predict the company's future risk, which is characterized by its volatility  $v_i$ . Such a prediction can be defined by a parameterized function  $f$  as follows:

$$\hat{v}_i = f(\mathbf{d}_i; \mathbf{w}). \quad (2)$$

The goal is to learn a  $p$ -dimensional vector  $\mathbf{w}$  from the training data  $T = \{(\mathbf{d}_i, v_i) | \mathbf{d}_i \in \mathbb{R}^p, v_i \in \mathbb{R}\}$ .

Support Vector Regression (SVR) (Drucker et al., 1997) is a popular technique for training such a regression model. SVR is trained by solving the following optimization problem:

$$\begin{aligned} \min_{\mathbf{w}} V(\mathbf{w}) &= \frac{1}{2} \langle \mathbf{w}, \mathbf{w} \rangle \\ &+ \frac{C}{n} \sum_{i=1}^n \max(|v_i - f(\mathbf{d}_i; \mathbf{w})| - \epsilon, 0), \end{aligned}$$

where  $C$  is a regularization constant and  $\epsilon$  controls the training error. More details about SVR can be found in (Schölkopf and Smola, 2001).

### 2.3.2 Ranking Task

For the ranking task, our goal is to rank companies by using their financial reports according to the volatilities of stock returns. Following the work in (Tsai and Wang, 2013), we split the volatilities of company stock returns within a year into different risk levels, which can be considered as the relative difference of risk among the companies.

After classifying the volatilities of stock returns (of companies) into different risk levels, the ranking task can be defined as follows: Given a collection of financial reports  $D$ , we aim to rank the companies via a ranking model  $f : \mathbb{R}^p \rightarrow \mathbb{R}$  such that the rank order of the set of companies is specified by the real value that the model  $f$  takes. In specific,  $f(\mathbf{d}_i) > f(\mathbf{d}_j)$  is taken to mean that the model asserts that  $c_i \succ c_j$ , where  $c_i \succ c_j$  means that  $c_i$  is ranked higher than  $c_j$ ; that is, the company  $c_i$  is more risky than  $c_j$ . In this paper, we adopt Ranking SVM (Joachims, 2006) for the ranking task.

## 3 Experiments

This section first describes the details of our experimental settings. Then, we report the experimental results of the models trained on the finance-specific

Year	# of Documents	# of Unique Terms
1996	1,406	19,613
1997	2,260	26,039
1998	2,461	29,020
1999	2,524	30,359
2000	2,424	30,312
2001	2,596	32,292
2002	2,845	38,692
2003	3,611	48,513
2004	3,558	50,674
2005	3,474	53,388
2006	3,306	51,147

Table 1: Statistics of the Corpora.

Dictionary	# of Words	# of Stemmed Words
Fin-Neg	2,349	918
Fin-Pos	354	151
Fin-Unc	291	127
Fin-Lit	871	443
MW-Strong	19	10
MW-Weak	27	15
Total	3,911	1,664

Table 2: Statistics of the Financial Lexicon.

sentiments only and those on original texts for the regression and ranking tasks.

### 3.1 Experimental Settings

#### 3.1.1 Corpora and Preprocessings

In the United States, the federal securities laws require publicly traded companies to disclose information on a regular basis. A Form 10-K, an annual report required by the Securities and Exchange Commission (SEC), provides a comprehensive overview of the company’s business and financial conditions, and includes audited financial statements. In this paper, the 10-K Corpus (Kogan et al., 2009) is used to conduct our experiments, in which only Section 7 “management’s discussion and analysis of financial conditions and results of operations” (MD&A) is used because the section contains the most important forward-looking statements about the companies.

For the preprocessing, in our experiments, all documents and the 6 financial sentiment word lists were stemmed by the Porter stemmer, and some stop words were also removed. Table 1 lists the statistics of documents and unique terms in each year. Table 2 shows the statistics before and after

stemming in each of the 6 financial word lists. Note that some words occur in more than one word list, so the number of unique stemmed sentiment words is 1,546 rather than 1,664.

In addition, the twelve months before/after the report volatility for each company (denote as  $v^{-(12)}$  and  $v^{+(12)}$ , respectively) can be calculated by Equation (1), where the price return series can be obtained from the Center for Research in Security Prices (CRSP) US Stocks Database. For the ranking task, in order to obtain the relative risks among companies, we categorize the companies of each year into 5 risk levels by following the work in (Tsai and Wang, 2013).

#### 3.1.2 Feature Representation

In our experiments, for the bag-of-words model, two word features are used to represent the 10-K reports. Given a document  $\mathbf{d}$ , two word features (i.e., TFIDF and LOG1P) are calculated as follows:

- $\text{TFIDF}(t, \mathbf{d}) = \text{TF}(t, \mathbf{d}) \times \text{IDF}(t, \mathbf{d}) = \text{TC}(t, \mathbf{d})/|\mathbf{d}| \times \log(|D|/|\mathbf{d} \in D : t \in \mathbf{d}|)$ ,
- $\text{LOG1P} = \log(1 + \text{TC}(t, \mathbf{d}))$ .

Above,  $\text{TC}(t, \mathbf{d})$  denotes the term count of  $t$  in  $\mathbf{d}$ ,  $|\mathbf{d}|$  is the length of document  $\mathbf{d}$ , and  $D$  denotes the set of all documents in each year. Note that IDF is computed from the documents in a single year because the document frequency of a specific word may vary across different years. Following (Kogan et al., 2009), we also use the logarithm of the twelve months before the report volatility (i.e.,  $\log v^{-(12)}$ ) as an additional feature. We denote these trained models as TFIDF+ and LOG1P+ hereafter.

#### 3.1.3 Evaluation Metrics

For the regression task, the performance is measured by the Mean Squared Error (MSE) between the predicted ( $\hat{v}_i^{+(12)}$ ) and true log-volatilities ( $v_i^{+(12)}$ ).

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n \left( \log \left( v_i^{+(12)} \right) - \log \left( \hat{v}_i^{+(12)} \right) \right)^2,$$

where  $n$  is the number of tested companies.

For the ranking task, two rank correlation metrics are used to evaluate the performance in our experiments: Spearman’s Rho (Myers and Well, 2003) and Kendall’s Tau (Kendall, 1938). Given two ranked lists  $X = \{x_1, x_2, \dots, x_n\}$  and  $Y =$

Task (Features)		2001	2002	2003	2004	2005	2006	Mirco-avg
Regression (LOG1P+)	Mean Squared Error							
	ORG	<b>0.18082</b>	0.17175	0.17157	0.12879	0.13038	0.14287	0.15271
	SEN	0.18506	<b>0.16367</b>	<b>0.15795</b>	<b>0.12822</b>	<b>0.13029</b>	<b>0.13998</b>	<b>0.14894</b>
Ranking (TFIDF+)	Kendall's Tau							
	ORG	0.62173	<b>0.63626</b>	0.58528	<b>0.59350</b>	0.59651	0.57641	0.59965
	SEN	<b>0.63349</b>	0.62280	<b>0.60527</b>	0.59017	<b>0.60273</b>	<b>0.58287</b>	<b>0.60458</b>
	Spearman's Rho							
	ORG	0.65271	<b>0.66692</b>	0.61662	<b>0.62317</b>	0.62531	0.60371	0.62939
	SEN	<b>0.66397</b>	0.65303	<b>0.63646</b>	0.61953	<b>0.63133</b>	<b>0.60999</b>	<b>0.63403</b>

Table 3: Experimental Results of Using Original Texts and Only Sentiment Words.

$\{y_1, y_2, \dots, y_n\}$ ,

$$\text{Rho} = 1 - \frac{6 \sum (x_i - y_i)^2}{n(n^2 - 1)},$$

$$\text{Tau} = \frac{\#\text{concordant pairs} - \#\text{discordant pairs}}{0.5 \cdot n \cdot (n - 1)}.$$

For the measure of Kendall's Tau, any pair of observations  $(x_i, y_i)$  and  $(x_j, y_j)$  is concordant if the ranks for both elements agree; that is, if both  $x_i \succ x_j$  and  $y_i \succ y_j$  or if both  $x_j \succ x_i$  and  $y_j \succ y_i$ . In contrast, it is discordant if  $x_i \succ x_j$  and  $y_j \succ y_i$  or if  $x_j \succ x_i$  and  $y_i \succ y_j$ . If  $x_i = x_j$  or  $y_i = y_j$ , the pair is neither concordant nor discordant.

### 3.1.4 Parameter Settings

For the regression task, linear kernel is adopted with  $\epsilon = 0.1$  and the trade-off  $C$  is set to the default value of SVM<sup>light</sup>,<sup>2</sup> which are the similar settings to those in (Kogan et al., 2009). For ranking, linear kernel is adopted with  $C = 1$ , all the other parameters are set as the default values of SVM<sup>Rank</sup>.<sup>3</sup>

## 3.2 Experimental Results

Table 3 tabulates the experimental results, in which the training data is composed of the financial reports in a five-year period, the following year of which is the test data. For example, the reports from year 1996 to 2000 constitute a training data, and the learned model is tested on the reports of year 2001.

We compare the performance of the models trained on the original texts (denoted as ORG hereafter) with those on only sentiment words (denoted

as SEN hereafter). In our experiments, the word feature LOG1P is chosen for the regression task and TFIDF for the ranking one, as suggested in (Kogan et al., 2009) and (Tsai and Wang, 2013). Note that in these two studies, their models are trained on the original texts and the results are listed in the row denoted as ORG in Table 3. The bold face number in the table denotes the best result between ORG and SEN. As shown in the table, for the two tasks, the results of using only sentiment words, in most cases, perform better than those of using the original texts.

## 4 Analysis

### 4.1 Ranking vs. Regression

Figure 1 shows the top 10 learned words from both the ranking (TFIDF+) and regression (LOGP+) models trained on sentiment words only (SEN); in addition, the figure also lists the accumulated numbers of these words appearing in the 6 corresponding regression or ranking models.

Observe that the words learned from the ranking models are much more consistent than those from the regression ones. For example, the words "amend," "deficit," "forbear" appear in all of the 6 ranking models; in addition, there are 7 words from the ranking models get the majority vote with more than 4 occurrences, whereas only 3 words from the regression ones occur more than 4 times. On the other hand, there are 11 words from the ranking models and 20 words from the regression ones that occur only one time. The results shown in Figure 1 correlate with the findings in (Tsai and Wang, 2013), which states that adopting the ranking models to analyze the relations between financial risk

<sup>2</sup><http://svmlight.joachims.org/>

<sup>3</sup>[http://www.cs.cornell.edu/people/tj/svm\\_light/svm\\_rank.html](http://www.cs.cornell.edu/people/tj/svm_light/svm_rank.html)

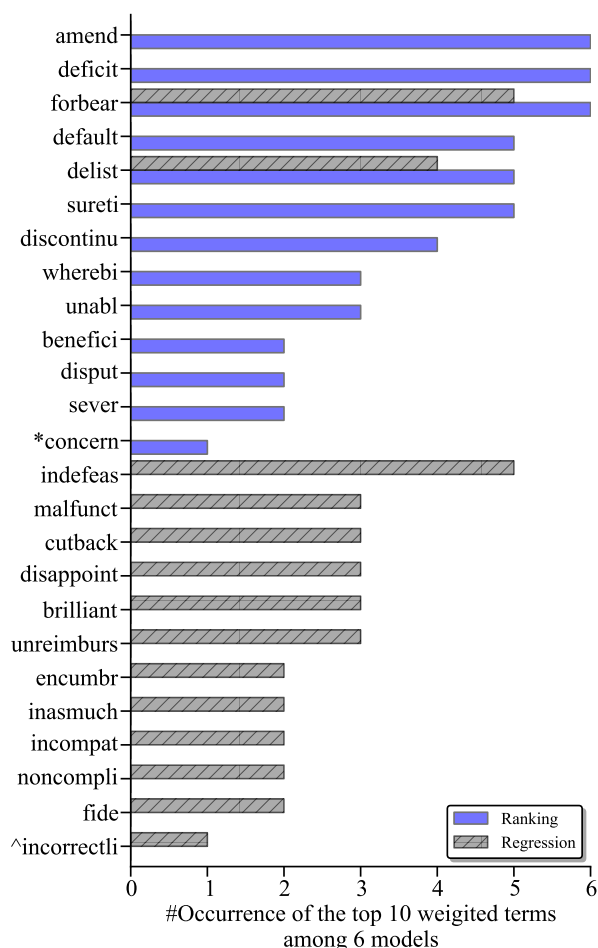


Figure 1: Number of Occurrences of the Top 10 Weighted Terms Learned via the Ranking and Regression Tasks. The notation \* denotes that except the term “concern” there are other terms that occur only one time among 6 ranking models, which are listed as follows: *breach, profit, violat, regain, uncomplet, accid, abl, integr, doubt, grantor*; similarly, for the notation ^, the terms are: *incorrectli, fault, nondisclosur, misus, breakag, defalc, excit, unclear, sentenc, overdu, omit, inforc, irrevoc, unencumb, further, variant, precipit, libel, loss*.

and text information might be a more reasonable way than the regression models.

## 4.2 Financial Sentiment Terms Analysis

As shown in Section 4.1, the ranking models can obtain more consistent results than the regression ones. Therefore, in the following discussions, we conduct some analyses on the words learned from the ranking models.

Figure 2 plots the words learned from our ranking models. In the figure, the single-outline circle denotes that only sentiment words are used as the training data; the double-outline circle denotes that all words in the original texts are considered when training. Moreover, the color filled in a circle with a term denotes which the sentiment word lists the

term belongs to; the circle with 2-mixed colors indicates the term belongs to two word lists. Note that the circle area is proportional to the average weight of each term.

In Figure 2, the top 5 average weighted words for the results of each kind of training data are marked by numbers from 1 to 5. For the case of training on sentiment words only (SEN), the top 5 average weighted words are *amend, deficit, forbear, delist, default*, whereas those under case ORG are *ceg, nasdaq, gnb, coven, forbear*; only one word *forbear* overlaps. An interesting finding is that when the models are trained on the original texts, some less informative terms like *ceg* (a company name, Co-Energy Group), *nasdaq* (an American stock exchange), *gnb* (a company name, GNB Technologies), are highly ranked; however, the relation is weak between these words and financial risk. In contrast, as only sentiment words are used for training, it is more reasonable that the terms are highly related to financial risk. In addition, since the terms in the figure have been stemmed, one term may correspond to one or more words. We also list the original words from the sentiment lexicon for each top 5 average weighted sentiment term in Figure 2. For example, the top 1 weighted term “*amend*” will have the list containing the words “*amend,*” “*amendable,*” “*amendatory,*” and so on.

Below we provide some original descriptions from 10-K reports that contain the top 2 weighted sentiment words in Figure 2. Note that the term with a higher weight is associated with higher financial risk. First, the term “*amend*” from the Fin-Lit list is considered. One piece of paragraph quoted from the original report is listed as follows:

(from AGO, 2006 Form 10-K)

On March 22, 2005, we *amended* the term loan agreements to, among other reasons, lower the borrowing rate by 25 basis points from LIBOR plus 2.00% to LIBOR plus 1.75%.

In finance, the *amend* usually means “to change by some formal processes.” This top-ranked term indicates that companies amending their policies frequently are associated with relative high risk.

We then discuss the term “*deficit*” from the Fin-Neg list, which means an excess of liabilities over assets, of losses over profits, or of expenditure over income in finance. Therefore, it is natural to say that a company associated with higher deficit might

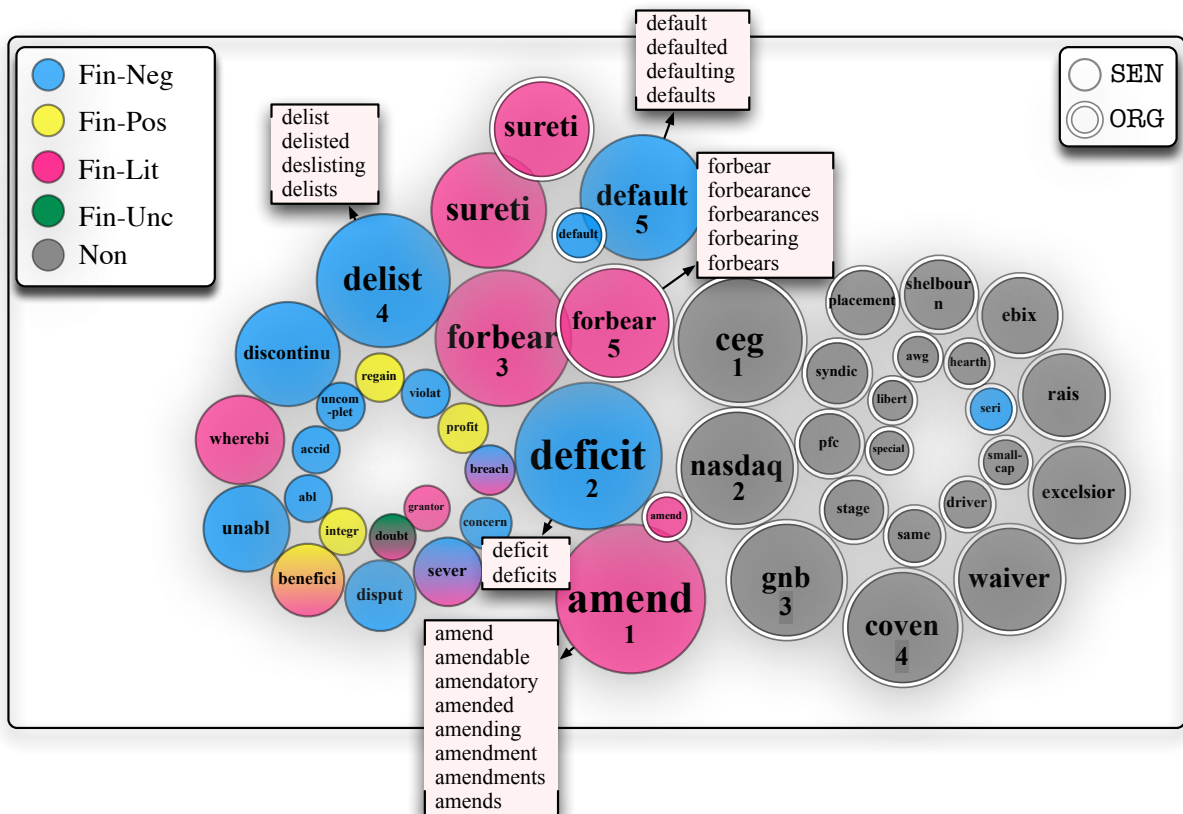


Figure 2: Highly-Weighted Terms Learned from the 6 Ranking Models of Using Original Texts (ORG) and Only Sentiment Words (SEN). The color filled in a circle with a term denotes which the sentiment word lists the word belongs to; the circle with 2-mixed colors indicates the term belongs to two word lists. The single-outline circle denotes that only sentiment words from the 6 dictionaries (see Table 2) are used as the training data; the double-outline circle denotes that the original texts are considered when training. Top 5 terms for the results of each kind of training data are marked by numbers from 1 to 5; the original words from the sentiment lexicon for each top 5 average weighted sentiment terms are also provided.

have higher risk. One piece of paragraph quoted from the original report is listed as follows:

(from AXS-One Inc., 2006 Form 10-K)  
 At December 31, 2005, we had cash and cash equivalents of \$3.6 million and a working capital *deficit* of \$3.6 million which included \$8.2 million of deferred revenue. The increase of the working capital *deficit* from \$3.3 million at December 31, 2004 is primarily the result of a decrease in cash and decreased accounts receivable offset partially by a decrease in deferred revenue.

## 5 Conclusions and Future Work

This paper identifies the importance of sentiment words in financial reports associated with financial risk. With the usage of a finance-specific sentiment lexicon, regression and ranking techniques are applied to analyze the relations between the sentiment words and financial risk. The experimental results

show that, based on the bag-of-words model, the models trained on sentiment words only can result in comparable performance to those on origin texts, which attests the importance of the financial sentiment words on risk prediction. In addition, the learned models also suggest strong correlations between financial sentiment words in financial reports and the risk of companies. As a result, these findings provide us more insight and understanding into the impact of financial sentiment words on companies' future risk. There are several future work, such as how to use even further information (i.e., syntactic information) for analysis, and how to conduct more fine-grained analysis.

## Acknowledgments

This research was partially supported by the National Science Council of Taiwan under the grants NSC 100-2218-E-133-001-MY2, 101-2221-E-004-017, 102-2221-E-004-006, and 102-2221-E-133-001-MY3.

## References

- H. Drucker, C.J.C. Burges, L. Kaufman, A. Smola, and V. Vapnik. 1997. Support vector regression machines. *Advances in neural information processing systems*, pages 155–161.
- Ronen Feldman. 2013. Techniques and applications for sentiment analysis. *Communications of the ACM*, 56(4):82–89.
- Diego Garca. 2013. Sentiment during recessions. *The Journal of Finance (Online Published)*.
- Thorsten Joachims. 2006. Training linear svms in linear time. In *KDD '06*, pages 217–226.
- M.G. Kendall. 1938. A new measure of rank correlation. *Biometrika*, 30(1/2):81–93.
- S. Kogan, D. Levin, B.R. Routledge, J.S. Sagi, and N.A. Smith. 2009. Predicting risk from financial reports with regression. In *NAACL '09*, pages 272–280.
- Tim Loughran and Bill McDonald. 2011. When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of Finance*, 66(1):35–65.
- Saif M. Mohammad and Peter D. Turney. 2010. Emotions evoked by common words and phrases: using mechanical turk to create an emotion lexicon. In *NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, pages 26–34.
- J. L. Myers and A. D. Well. 2003. *Research design and statistical analysis*.
- Ramanathan Narayanan, Bing Liu, and Alok Choudhary. 2009. Sentiment analysis of conditional sentences. In *EMNLP '09*, pages 180–189.
- S McKay Price, James S Doran, David R Peterson, and Barbara A Bliss. 2012. Earnings conference calls and stock returns: The incremental informativeness of textual tone. *Journal of Banking & Finance*, 36(4):992–1011.
- B. Schölkopf and A.J. Smola. 2001. *Learning with kernels: Support vector machines, regularization, optimization, and beyond*.
- Ming-Feng Tsai and Chuan-Ju Wang. 2013. Risk ranking from financial reports. In *Advances in Information Retrieval*, pages 804–807.
- R.S. Tsay. 2005. *Analysis of financial time series*.